1. Preface
2. Electric Charge and Electric Field
    1. Introduction to Electric Charge and Electric Field
    2. Static Electricity and Charge: Conservation of Charge
    3. Conductors and Insulators
    4. Coulomb’s Law
    5. Electric Field: Concept of a Field Revisited
    6. Electric Field Lines: Multiple Charges
    7. Electric Forces in Biology
    8. Conductors and Electric Fields in Static Equilibrium
    9. Applications of Electrostatics
3. Electric Potential and Electric Field
    1. Introduction to Electric Potential and Electric Energy
    2. Electric Potential Energy: Potential Difference
    3. Electric Potential in a Uniform Electric Field
    4. Electrical Potential Due to a Point Charge
    5. Equipotential Lines
    6. Capacitors and Dielectrics
    7. Capacitors in Series and Parallel
    8. Energy Stored in Capacitors
4. Electric Current, Resistance, and Ohm's Law
    1. Introduction to Electric Current, Resistance, and Ohm's Law
    2. Current
    3. Ohm’s Law: Resistance and Simple Circuits
    4. Resistance and Resistivity
    5. Electric Power and Energy
    6. Alternating Current versus Direct Current
    7. Electric Hazards and the Human Body
    8. Nerve Conduction–Electrocardiograms
5. Circuits, Bioelectricity, and DC Instruments

    1. Introduction to Circuits and DC Instruments
    2. Resistors in Series and Parallel
    3. Electromotive Force: Terminal Voltage
    4. Kirchhoff’s Rules
    5. DC Voltmeters and Ammeters
    6. Null Measurements
    7. DC Circuits Containing Resistors and Capacitors
6. Magnetism
    1. Introduction to Magnetism
    2. Magnets
    3. Ferromagnets and Electromagnets
    4. Magnetic Fields and Magnetic Field Lines
    5. Magnetic Field Strength: Force on a Moving Charge in a Magnetic Field
    6. Force on a Moving Charge in a Magnetic Field: Examples and Applications
    7. The Hall Effect
    8. Magnetic Force on a Current-Carrying Conductor
    9. Torque on a Current Loop: Motors and Meters
    10. Magnetic Fields Produced by Currents: Ampere’s Law
    11. Magnetic Force between Two Parallel Conductors
    12. More Applications of Magnetism
7. Electromagnetic Induction, AC Circuits, and Electrical Technologies
    1. Introduction to Electromagnetic Induction, AC Circuits and Electrical Technologies
    2. Induced Emf and Magnetic Flux
    3. Faraday’s Law of Induction: Lenz’s Law
    4. Motional Emf
    5. Eddy Currents and Magnetic Damping
    6. Electric Generators
    7. Back Emf

    8. Transformers
    9. Electrical Safety: Systems and Devices
    10. Inductance
    11. RL Circuits
    12. Reactance, Inductive and Capacitive
    13. RLC Series AC Circuits
8. Temperature, Kinetic Theory, and the Gas Laws
    1. Introduction to Temperature, Kinetic Theory, and the Gas Laws
    2. Temperature
    3. Thermal Expansion of Solids and Liquids
    4. The Ideal Gas Law
    5. Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature
    6. Phase Changes
    7. Humidity, Evaporation, and Boiling
9. Heat and Heat Transfer Methods
    1. Introduction to Heat and Heat Transfer Methods
    2. Heat
    3. Temperature Change and Heat Capacity
    4. Phase Change and Latent Heat
    5. Heat Transfer Methods
    6. Conduction
    7. Convection
    8. Radiation
10. Thermodynamics
    1. Introduction to Thermodynamics
    2. The First Law of Thermodynamics
    3. The First Law of Thermodynamics and Some Simple Processes
    4. Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency

    5. Carnot’s Perfect Heat Engine: The Second Law of Thermodynamics Restated
    6. Applications of Thermodynamics: Heat Pumps and Refrigerators
    7. Entropy and the Second Law of Thermodynamics: Disorder and the Unavailability of Energy
    8. Statistical Interpretation of Entropy and the Second Law of Thermodynamics: The Underlying Explanation
11. Fluid Statics
    1. Introduction to Fluid Statics
    2. What Is a Fluid?
    3. Density
    4. Pressure
    5. Variation of Pressure with Depth in a Fluid
    6. Pascal’s Principle
    7. Gauge Pressure, Absolute Pressure, and Pressure Measurement
    8. Archimedes’ Principle
    9. Cohesion and Adhesion in Liquids: Surface Tension and Capillary Action
    10. Pressures in the Body
12. Fluid Dynamics and Its Biological and Medical Applications
    1. Introduction to Fluid Dynamics and Its Biological and Medical Applications
    2. Flow Rate and Its Relation to Velocity
    3. Bernoulli’s Equation
    4. The Most General Applications of Bernoulli’s Equation
    5. Viscosity and Laminar Flow; Poiseuille’s Law
    6. The Onset of Turbulence
    7. Motion of an Object in a Viscous Fluid
    8. Molecular Transport Phenomena: Diffusion, Osmosis, and Related Processes

13. Physics of Hearing
    1. Introduction to the Physics of Hearing
    2. Sound
    3. Speed of Sound, Frequency, and Wavelength
    4. Sound Intensity and Sound Level
    5. Doppler Effect and Sonic Booms
    6. Sound Interference and Resonance: Standing Waves in Air Columns
    7. Hearing
    8. Ultrasound
14. Atomic Masses
15. Useful Information
16. Glossary of Key Symbols and Notation

Preface

Welcome to *College Physics*, an OpenStax resource. This textbook was written to increase student access to high-quality learning materials, maintaining highest standards of academic rigor at little to no cost.

About OpenStax

OpenStax is a nonprofit based at Rice University, and it’s our mission to improve student access to education. Our first openly licensed college textbook was published in 2012, and our library has since scaled to over 20 books for college and AP courses used by hundreds of thousands of students. Our adaptive learning technology, designed to improve learning outcomes through personalized educational paths, is being piloted in college courses throughout the country. Through our partnerships with philanthropic foundations and our alliance with other educational resource organizations, OpenStax is breaking down the most common barriers to learning and empowering students and instructors to succeed.

About OpenStax Resources

Customization

College Physics is licensed under a Creative Commons Attribution 4.0 International (CC BY) license, which means that you can distribute, remix, and build upon the content, as long as you provide attribution to OpenStax and its content contributors.

Because our books are openly licensed, you are free to use the entire book or pick and choose the sections that are most relevant to the needs of your course. Feel free to remix the content by assigning your students certain chapters and sections in your syllabus, in the order that you prefer. You can even provide a direct link in your syllabus to the sections in the web view of your book.

Instructors also have the option of creating a customized version of their OpenStax book. The custom version can be made available to students in low-cost print or digital form through their campus bookstore. Visit your book page on openstax.org for more information.

Errata

All OpenStax textbooks undergo a rigorous review process. However, like any professional-grade textbook, errors sometimes occur. Since our books are web based, we can make updates periodically when deemed pedagogically necessary. If you have a correction to suggest, submit it through the link on your book page on openstax.org. Subject matter experts review all errata suggestions. OpenStax is committed to remaining transparent about all updates, so you will also find a list of past errata changes on your book page on openstax.org.

Format

You can access this textbook for free in web view or PDF through openstax.org, and in low-cost print and iBooks editions.

About *College Physics*

College Physics meets standard scope and sequence requirements for a two-semester introductory algebra-based physics course. The text is grounded in real-world examples to help students grasp fundamental physics concepts. It requires knowledge of algebra and some trigonometry, but not calculus. *College Physics* includes learning objectives, concept questions, links to labs and simulations, and ample practice opportunities for traditional physics application problems.

Coverage and Scope

College Physics is organized such that topics are introduced conceptually with a steady progression to precise definitions and analytical applications. The analytical aspect (problem solving) is tied back to the conceptual before moving on to another topic. Each introductory chapter, for example, opens with an engaging photograph relevant to the subject of the chapter and interesting applications that are easy for most students to visualize.

Chapter 1: Introduction: The Nature of Science and Physics
Chapter 2: Kinematics
Chapter 3: Two-Dimensional Kinematics
Chapter 4: Dynamics: Force and Newton's Laws of Motion
Chapter 5: Further Applications of Newton's Laws: Friction, Drag, and Elasticity
Chapter 6: Uniform Circular Motion and Gravitation
Chapter 7: Work, Energy, and Energy Resources
Chapter 8: Linear Momentum and Collisions
Chapter 9: Statics and Torque
Chapter 10: Rotational Motion and Angular Momentum
Chapter 11: Fluid Statics
Chapter 12: Fluid Dynamics and Its Biological and Medical Applications
Chapter 13: Temperature, Kinetic Theory, and the Gas Laws
Chapter 14: Heat and Heat Transfer Methods
Chapter 15: Thermodynamics
Chapter 16: Oscillatory Motion and Waves
Chapter 17: Physics of Hearing
Chapter 18: Electric Charge and Electric Field
Chapter 19: Electric Potential and Electric Field
Chapter 20: Electric Current, Resistance, and Ohm's Law
Chapter 21: Circuits and DC Instruments
Chapter 22: Magnetism
Chapter 23: Electromagnetic Induction, AC Circuits, and Electrical Technologies
Chapter 24: Electromagnetic Waves
Chapter 25: Geometric Optics
Chapter 26: Vision and Optical Instruments
Chapter 27: Wave Optics
Chapter 28: Special Relativity

Chapter 29: Introduction to Quantum Physics
Chapter 30: Atomic Physics
Chapter 31: Radioactivity and Nuclear Physics
Chapter 32: Medical Applications of Nuclear Physics
Chapter 33: Particle Physics
Chapter 34: Frontiers of Physics
Appendix A: Atomic Masses
Appendix B: Selected Radioactive Isotopes
Appendix C: Useful Information
Appendix D: Glossary of Key Symbols and Notation

Concepts and Calculations

The ability to calculate does not guarantee conceptual understanding. In order to unify conceptual, analytical, and calculation skills within the learning process, we have integrated Strategies and Discussions throughout the text.

Modern Perspective

The chapters on modern physics are more complete than many other texts on the market, with an entire chapter devoted to medical applications of nuclear physics and another to particle physics. The final chapter of the text, “Frontiers of Physics,” is devoted to the most exciting endeavors in physics. It ends with a module titled “Some Questions We Know to Ask.”

Key Features

Modularity

This textbook is organized as a collection of modules that can be rearranged and modified to suit the needs of a particular professor or class. That being said, modules often contain references to content in other modules, as most topics in physics cannot be discussed in isolation.

Learning Objectives

Every module begins with a set of learning objectives. These objectives are designed to guide the instructor in deciding what content to include or assign, and to guide the student with respect to what he or she can expect to learn. After completing the module and end-of-module exercises, students should be able to demonstrate mastery of the learning objectives.

Call-Outs

Key definitions, concepts, and equations are called out with a special design treatment. Call-outs are designed to catch readers’ attention, to make it clear that a specific term, concept, or equation is particularly important, and to provide easy reference for a student reviewing content.

Key Terms

Key terms are in bold and are followed by a definition in context. Definitions of key terms are also listed in the Glossary, which appears at the end of the module.

Worked Examples

Worked examples have four distinct parts to promote both analytical and conceptual skills. Worked examples are introduced in words, always using some application that should be of interest. This is followed by a Strategy section that emphasizes the concepts involved and how solving the problem relates to those concepts. This is followed by the mathematical Solution and Discussion.

Many worked examples contain multiple-part problems to help the students learn how to approach normal situations, in which problems tend to have multiple parts. Finally, worked examples employ the techniques of the problem-solving strategies so that students can see how those strategies succeed in practice as well as in theory.

Problem-Solving Strategies

Problem-solving strategies are first presented in a special section and subsequently appear at crucial points in the text where students can benefit most from them. Problem-solving strategies have a logical structure that is reinforced in the worked examples and supported in certain places by line drawings that illustrate various steps.

Misconception Alerts

Students come to physics with preconceptions from everyday experiences and from previous courses. Some of these preconceptions are misconceptions, and many are very common among students and the general public. Some are inadvertently picked up through misunderstandings of lectures and texts. The Misconception Alerts feature is designed to point these out and correct them explicitly.

Take-Home Investigations

Take Home Investigations provide the opportunity for students to apply or explore what they have learned with a hands-on activity.

Things Great and Small

In these special topic essays, macroscopic phenomena (such as air pressure) are explained with submicroscopic phenomena (such as atoms bouncing off walls). These essays support the modern perspective by describing aspects of modern physics before they are formally treated in later chapters. Connections are also made between apparently disparate phenomena.

Simulations

Where applicable, students are directed to the interactive PHeT physics simulations developed by the University of Colorado. There they can further explore the physics concepts they have learned about in the module.

Summary

Module summaries are thorough and functional and present all important definitions and equations. Students are able to find the definitions of all terms and symbols as well as their physical relationships. The structure of the summary makes plain the fundamental principles of the module or collection and serves as a useful study guide.

Glossary

At the end of every module or chapter is a Glossary containing definitions of all of the key terms in the module or chapter.

End-of-Module Problems

At the end of every chapter is a set of Conceptual Questions and/or skills-based Problems & Exercises. Conceptual Questions challenge students' ability to explain what they have learned conceptually, independent of the mathematical details. Problems & Exercises challenge students to apply both concepts and skills to solve mathematical physics problems. Online, every other problem includes an answer that students can reveal immediately by clicking on a “Show Solution” button.

In addition to traditional skills-based problems, there are three special types of end-of-module problems: Integrated Concept Problems, Unreasonable Results Problems, and Construct Your Own Problems. All of these problems are indicated with a subtitle preceding the problem.

Integrated Concept Problems

In Integrated Concept Problems, students are asked to apply what they have learned about two or more concepts to arrive at a solution to a problem. These problems require a higher level of thinking because, before solving a problem, students have to recognize the combination of strategies required to solve it.

Unreasonable Results

In Unreasonable Results Problems, students are challenged to not only apply concepts and skills to solve a problem, but also to analyze the answer with respect to how likely or realistic it really is. These problems contain a premise that produces an unreasonable answer and are designed to further emphasize that properly applied physics must describe nature accurately and is not simply the process of solving equations.

Construct Your Own Problem

These problems require students to construct the details of a problem, justify their starting assumptions, show specific steps in the problem’s solution, and finally discuss the meaning of the result. These types of problems relate well to both conceptual and analytical aspects of physics, emphasizing that physics must describe nature. Often they involve an integration of topics from more than one chapter. Unlike other problems, solutions are not provided since there is no single correct answer.

Instructors should feel free to direct students regarding the level and scope of their considerations. Whether the problem is solved and described correctly will depend on initial assumptions.

Additional Resources

Student and Instructor Resources

We’ve compiled additional resources for both students and instructors, including Getting Started Guides, an instructor solution manual, and PowerPoint slides. Instructor resources require a verified instructor account, which can be requested on your openstax.org log-in. Take advantage of these resources to supplement your OpenStax book.

Partner Resources

OpenStax Partners are our allies in the mission to make high-quality learning materials affordable and accessible to students and instructors everywhere. Their tools integrate seamlessly with our OpenStax titles at a low cost. To access the partner resources for your text, visit your book page on openstax.org.

About the Authors

Senior Contributing Authors

Paul Peter Urone, Professor Emeritus at California State University, Sacramento
Roger Hinrichs, State University of New York, College at Oswego

Contributing Authors

Kim Dirks, University of Auckland
Manjula Sharma, University of Sydney

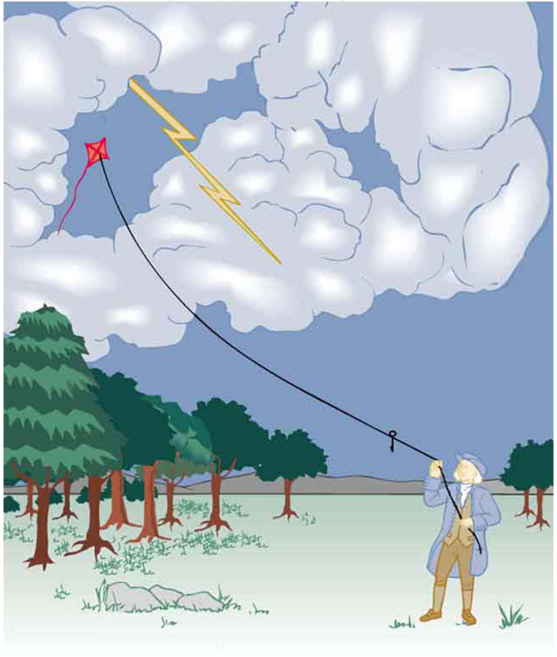
Reviewers

Matthew Adams, Crafton Hills College, San Bernardino Community College District
Erik Christensen, South Florida Community College
Douglas Ingram, Texas Christian University
Eric Kincanon, Gonzaga University
Lee H. LaRue, Paris Junior College
Chuck Pearson, Virginia Intermont College
Marc Sher, College of William and Mary
Ulrich Zurcher, Cleveland State University

Introduction to Electric Charge and Electric Field class="introduction"

Static electricity from this plastic slide causes the child’s hair to stand on end. The sliding motion stripped electrons away from the child’s body, leaving an excess of positive charges, which repel each other along each strand of hair. (credit: Ken Bosma/Wikimedia Commons)

The image of American politician and scientist Benjamin Franklin (1706–1790) flying a kite in a thunderstorm is familiar to every schoolchild. (See [link].) In this experiment, Franklin demonstrated a connection between lightning and **static electricity**. Sparks were drawn from a key hung on a kite string during an electrical storm. These sparks were like those produced by static electricity, such as the spark that jumps from your finger to a metal doorknob after you walk across a wool carpet. What Franklin demonstrated in his dangerous experiment was a connection between phenomena on two different scales: one the grand power of an electrical storm, the other an effect of more human proportions. Connections like this one reveal the underlying unity of the laws of nature, an aspect we humans find particularly appealing.



When Benjamin Franklin demonstrated that lightning was related to static electricity, he made a connection that is now part of the evidence that all directly experienced forces except the gravitational force are manifestations of the electromagnetic force.

Much has been written about Franklin. His experiments were only part of the life of a man who was a scientist, inventor, revolutionary, statesman, and writer. Franklin’s experiments were not performed in isolation, nor were they the only ones to reveal connections.

For example, the Italian scientist Luigi Galvani (1737–1798) performed a series of experiments in which static electricity was used to stimulate contractions of leg muscles of dead frogs, an effect already known in humans subjected to static discharges. But Galvani also found that if he joined two metal wires (say copper and zinc) end to end and touched the other ends to muscles, he produced the same effect in frogs as static discharge. Alessandro Volta (1745–1827), partly inspired by Galvani’s work, experimented with various combinations of metals and developed the battery.

During the same era, other scientists made progress in discovering fundamental connections. The periodic table was developed as the systematic properties of the elements were discovered. This influenced the development and refinement of the concept of atoms as the basis of matter. Such submicroscopic descriptions of matter also help explain a great deal more.

Atomic and molecular interactions, such as the forces of friction, cohesion, and adhesion, are now known to be manifestations of the **electromagnetic force**. Static electricity is just one aspect of the electromagnetic force, which also includes moving electricity and magnetism.

All the macroscopic forces that we experience directly, such as the sensations of touch and the tension in a rope, are due to the electromagnetic force, one of the four fundamental forces in nature. The gravitational force, another fundamental force, is actually sensed through the electromagnetic interaction of molecules, such as between those in our feet and those on the top of a bathroom scale. (The other two fundamental forces, the strong nuclear force and the weak nuclear force, cannot be sensed on the human scale.)

This chapter begins the study of electromagnetic phenomena at a fundamental level. The next several chapters will cover static electricity, moving electricity, and magnetism—collectively known as electromagnetism. In this chapter, we begin with the study of electric phenomena due to charges that are at least temporarily stationary, called electrostatics, or static electricity.

Glossary

static electricity
: a buildup of electric charge on the surface of an object

electromagnetic force
: one of the four fundamental forces of nature; the electromagnetic force consists of static electricity, moving electricity and magnetism

Static Electricity and Charge: Conservation of Charge

- Define electric charge, and describe how the two types of charge interact.
- Describe three common situations that generate static electricity.
- State the law of conservation of charge.

Borneo amber was mined in Sabah, Malaysia, from shale-sandstone-mudstone veins. When a piece of amber is rubbed with a piece of silk, the amber gains more electrons, giving it a net negative charge. At the same time, the silk, having lost electrons, becomes positively charged. (credit: Sebakoamber, Wikimedia Commons)

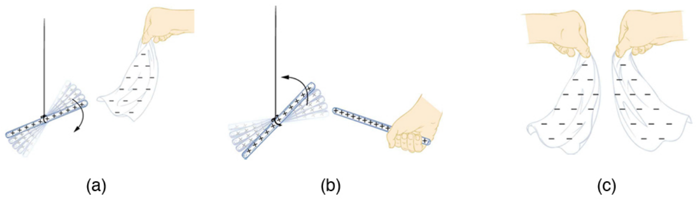
What makes plastic wrap cling? Static electricity. Not only are applications of static electricity common these days, its existence has been known since ancient times. The first record of its effects dates to ancient Greeks who noted more than 500 years B.C. that polishing amber temporarily enabled it to attract bits of straw (see [link]). The very word *electric* derives from the Greek word for amber (*electron*).

Many of the characteristics of static electricity can be explored by rubbing things together. Rubbing creates the spark you get from walking across a wool carpet, for example. Static cling generated in a clothes dryer and the attraction of straw to recently polished amber also result from rubbing. Similarly, lightning results from air movements under certain weather conditions. You can also rub a balloon on your hair, and the static electricity created can then make the balloon cling to a wall. We also have to be cautious of static electricity, especially in dry climates. When we pump gasoline, we are warned to discharge ourselves (after sliding across the seat) on a metal surface before grabbing the gas nozzle. Attendants in hospital operating rooms must wear booties with aluminum foil on the bottoms to avoid creating sparks which may ignite the oxygen being used.

Some of the most basic characteristics of static electricity include:

- The effects of static electricity are explained by a physical quantity not previously introduced, called electric charge.
- There are only two types of charge, one called positive and the other called negative.
- Like charges repel, whereas unlike charges attract.
- The force between charges decreases with distance.

How do we know there are two types of **electric charge**? When various materials are rubbed together in controlled ways, certain combinations of materials always produce one type of charge on one material and the opposite type on the other. By convention, we call one type of charge "positive", and the other type "negative." For example, when glass is rubbed with silk, the glass becomes positively charged and the silk negatively charged. Since the glass and silk have opposite charges, they attract one another like clothes that have rubbed together in a dryer. Two glass rods rubbed with silk in this manner will repel one another, since each rod has positive charge on it. Similarly, two silk cloths so rubbed will repel, since both cloths have negative charge. [link] shows how these simple materials can be used to explore the nature of the force between charges.



A glass rod becomes positively charged when rubbed with silk, while the silk becomes negatively charged. (a) The glass rod is attracted to the silk because their charges are opposite. (b) Two similarly charged glass rods repel. (c) Two similarly charged silk cloths repel.

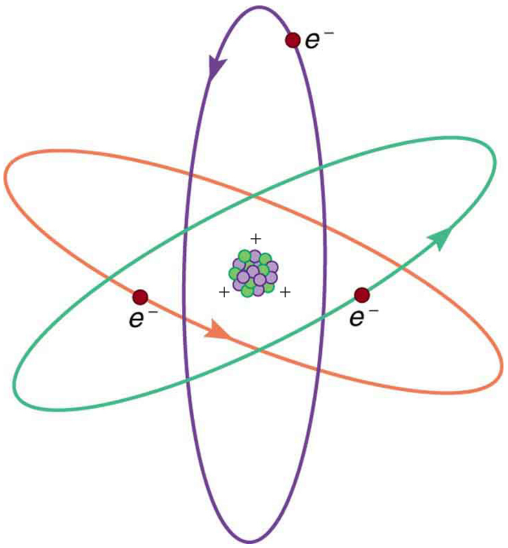
More sophisticated questions arise. Where do these charges come from? Can you create or destroy charge? Is there a smallest unit of charge? Exactly how does the force depend on the amount of charge and the distance between charges? Such questions obviously occurred to Benjamin Franklin and other early researchers, and they interest us even today.

Charge Carried by Electrons and Protons

Franklin wrote in his letters and books that he could see the effects of electric charge but did not understand what caused the phenomenon. Today we have the advantage of knowing that normal matter is made of atoms, and that atoms contain positive and negative charges, usually in equal amounts.

[link] shows a simple model of an atom with negative **electrons** orbiting its positive nucleus. The nucleus is positive due to the presence of positively charged **protons**. Nearly all charge in nature is due to electrons and protons, which are two of the three building blocks of most matter. (The third is the neutron, which is neutral, carrying no charge.) Other charge-carrying particles are observed in cosmic rays and nuclear decay, and are created in

particle accelerators. All but the electron and proton survive only a short time and are quite rare by comparison.



This simplified (and not to scale) view of an atom is called the planetary model of the atom. Negative electrons orbit a much heavier positive nucleus, as the planets orbit the much heavier sun. There the similarity ends, because forces in the atom are electromagnetic, whereas those in the planetary system are gravitational. Normal macroscopic amounts of matter contain immense numbers of atoms and molecules and, hence, even greater numbers of individual

negative and positive
charges.

The charges of electrons and protons are identical in magnitude but opposite in sign. Furthermore, all charged objects in nature are integral multiples of this basic quantity of charge, meaning that all charges are made of combinations of a basic unit of charge. Usually, charges are formed by combinations of electrons and protons. The magnitude of this basic charge is

Equation:

$$\mid q_e \mid = 1.60 \times 10^{-19}\ \text{C}.$$

The symbol q is commonly used for charge and the subscript e indicates the charge of a single electron (or proton).

The SI unit of charge is the coulomb (C). The number of protons needed to make a charge of 1.00 C is

Equation:

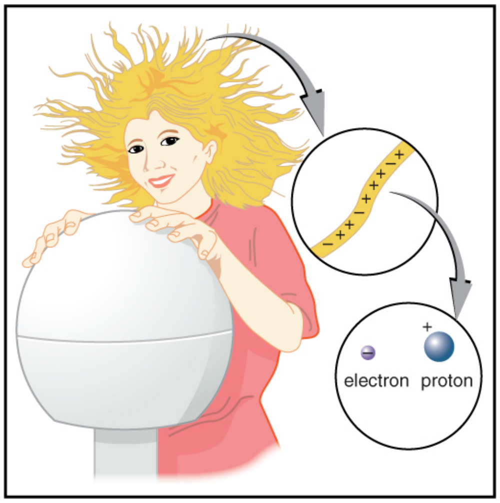
$$1.00\ \text{C} \times \frac{1\ \text{proton}}{1.60 \times 10^{-19}\ \text{C}} = 6.25 \times 10^{18}\ \text{protons}.$$

Similarly, 6.25×10^{18} electrons have a combined charge of −1.00 coulomb. Just as there is a smallest bit of an element (an atom), there is a smallest bit of charge. There is no directly observed charge smaller than $\mid q_e \mid$ (see Things Great and Small: The Submicroscopic Origin of Charge), and all observed charges are integral multiples of $\mid q_e \mid$.

Note:
Things Great and Small: The Submicroscopic Origin of Charge

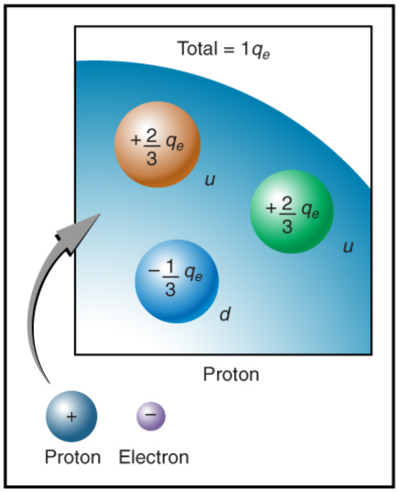
With the exception of exotic, short-lived particles, all charge in nature is carried by electrons and protons. Electrons carry the charge we have named negative. Protons carry an equal-magnitude charge that we call positive. (See [link].) Electron and proton charges are considered fundamental building blocks, since all other charges are integral multiples of those carried by electrons and protons. Electrons and protons are also two of the three fundamental building blocks of ordinary matter. The neutron is the third and has zero total charge.

[link] shows a person touching a Van de Graaff generator and receiving excess positive charge. The expanded view of a hair shows the existence of both types of charges but an excess of positive. The repulsion of these positive like charges causes the strands of hair to repel other strands of hair and to stand up. The further blowup shows an artist's conception of an electron and a proton perhaps found in an atom in a strand of hair.



When this person touches a Van de Graaff generator, she receives an excess of positive charge, causing her hair to stand on end. The charges in one hair are shown. An artist’s conception of an electron and a proton illustrate the particles carrying the negative and positive charges. We cannot really see these particles with visible light because they are so small (the electron seems to be an infinitesimal point), but we know a great deal about their measurable properties, such as the charges they carry.

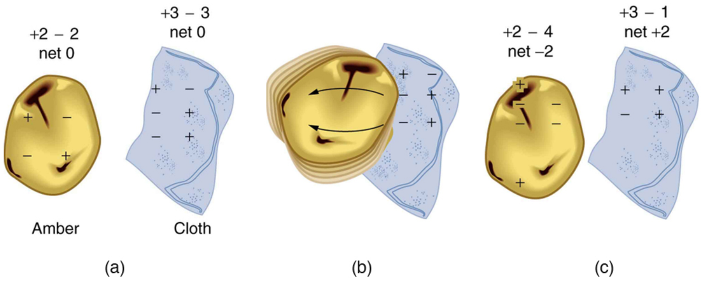
The electron seems to have no substructure; in contrast, when the substructure of protons is explored by scattering extremely energetic electrons from them, it appears that there are point-like particles inside the proton. These sub-particles, named quarks, have never been directly observed, but they are believed to carry fractional charges as seen in [link]. Charges on electrons and protons and all other directly observable particles are unitary, but these quark substructures carry charges of either $-\frac{1}{3}$ or $+\frac{2}{3}$. There are continuing attempts to observe fractional charge directly and to learn of the properties of quarks, which are perhaps the ultimate substructure of matter.



Artist’s conception of fractional quark charges inside a proton. A group of three quark charges add up to the single positive charge on the proton: $-\frac{1}{3}q_e + \frac{2}{3}q_e + \frac{2}{3}q_e = +1q_e$.

Separation of Charge in Atoms

Charges in atoms and molecules can be separated—for example, by rubbing materials together. Some atoms and molecules have a greater affinity for electrons than others and will become negatively charged by close contact in rubbing, leaving the other material positively charged. (See [link].) Positive charge can similarly be induced by rubbing. Methods other than rubbing can also separate charges. Batteries, for example, use combinations of substances that interact in such a way as to separate charges. Chemical interactions may transfer negative charge from one substance to the other, making one battery terminal negative and leaving the first one positive.



When materials are rubbed together, charges can be separated, particularly if one material has a greater affinity for electrons than another. (a) Both the amber and cloth are originally neutral, with equal positive and negative charges. Only a tiny fraction of the charges are involved, and only a few of them are shown here. (b) When rubbed together, some negative charge is transferred to the amber, leaving the cloth with a net positive charge. (c) When separated, the amber and cloth now have net charges, but the absolute value of the net positive and negative charges will be equal.

No charge is actually created or destroyed when charges are separated as we have been discussing. Rather, existing charges are moved about. In fact, in all situations the total amount of charge is always constant. This universally obeyed law of nature is called the **law of conservation of charge**.

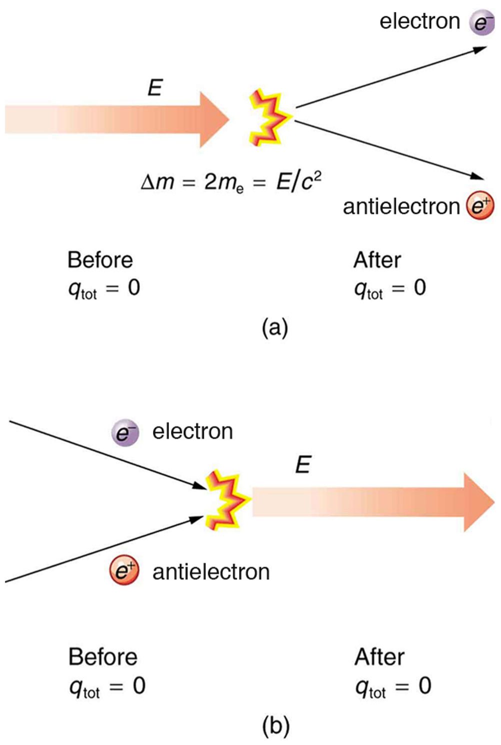
Note:
Law of Conservation of Charge
Total charge is constant in any process.

In more exotic situations, such as in particle accelerators, mass, Δm, can be created from energy in the amount $\Delta m = \frac{E}{c^2}$. Sometimes, the created mass is charged, such as when an electron is created. Whenever a charged particle is created, another having an opposite charge is always created along with it, so that the total charge created is zero. Usually, the two particles are “matter-antimatter” counterparts. For example, an antielectron would usually be created at the same time as an electron. The antielectron has a positive charge (it is called a positron), and so the total charge created is zero. (See [link].) All particles have antimatter counterparts with opposite signs. When matter and antimatter counterparts are brought together, they completely annihilate one another. By annihilate, we mean that the mass of the two particles is converted to energy E, again obeying the relationship $\Delta m = \frac{E}{c^2}$. Since the two particles have equal and opposite charge, the total charge is zero before and after the annihilation; thus, total charge is conserved.

Note:

Making Connections: Conservation Laws

Only a limited number of physical quantities are universally conserved. Charge is one—energy, momentum, and angular momentum are others. Because they are conserved, these physical quantities are used to explain more phenomena and form more connections than other, less basic quantities. We find that conserved quantities give us great insight into the rules followed by nature and hints to the organization of nature. Discoveries of conservation laws have led to further discoveries, such as the weak nuclear force and the quark substructure of protons and other particles.


electron
e^-
E
$\Delta m = 2m_e = E/c^2$
antielectron
e^+
Before
$q_{tot} = 0$
After
$q_{tot} = 0$
(a)
electron
e^-
E
antielectron
e^+
Before
$q_{tot} = 0$
After
$q_{tot} = 0$
(b)

(a) When enough energy is present, it can be converted into matter. Here the matter created is an electron–antielectron pair. (m_e is the electron’s mass.) The total charge before and after this event is zero. (b) When matter and antimatter collide, they annihilate each other; the total charge is conserved at zero before and after the annihilation.

The law of conservation of charge is absolute—it has never been observed to be violated. Charge, then, is a special physical quantity, joining a very short list of other quantities in nature that are always conserved. Other conserved quantities include energy, momentum, and angular momentum.

Note:
PhET Explorations: Balloons and Static Electricity
Why does a balloon stick to your sweater? Rub a balloon on a sweater, then let go of the balloon and it flies over and sticks to the sweater. View the charges in the sweater, balloons, and the wall.
https://phet.colorado.edu/sims/html/balloons-and-static-electricity/latest/balloons-and-static-electricity_en.html

Section Summary

- There are only two types of charge, which we call positive and negative.
- Like charges repel, unlike charges attract, and the force between charges decreases with the square of the distance.
- The vast majority of positive charge in nature is carried by protons, while the vast majority of negative charge is carried by electrons.
- The electric charge of one electron is equal in magnitude and opposite in sign to the charge of one proton.
- An ion is an atom or molecule that has nonzero total charge due to having unequal numbers of electrons and protons.
- The SI unit for charge is the coulomb (C), with protons and electrons having charges of opposite sign but equal magnitude; the magnitude of this basic charge $\mid q_e \mid$ is
  Equation:
  $$\mid q_e \mid = 1.60 \times 10^{-19}\text{ C}.$$
- Whenever charge is created or destroyed, equal amounts of positive and negative are involved.
- Most often, existing charges are separated from neutral objects to obtain some net charge.

- Both positive and negative charges exist in neutral objects and can be separated by rubbing one object with another. For macroscopic objects, negatively charged means an excess of electrons and positively charged means a depletion of electrons.
- The law of conservation of charge ensures that whenever a charge is created, an equal charge of the opposite sign is created at the same time.

Conceptual Questions

Exercise:

Problem:

There are very large numbers of charged particles in most objects. Why, then, don’t most objects exhibit static electricity?

Exercise:

Problem:

Why do most objects tend to contain nearly equal numbers of positive and negative charges?

Problems & Exercises

Exercise:

Problem:

Common static electricity involves charges ranging from nanocoulombs to microcoulombs. (a) How many electrons are needed to form a charge of -2.00 nC (b) How many electrons must be removed from a neutral object to leave a net charge of $0.500 \ \mu\text{C}$?

Solution:

(a) 1.25×10^{10}

(b) 3.13×10^{12}

Exercise:

Problem:

If 1.80×10^{20} electrons move through a pocket calculator during a full day’s operation, how many coulombs of charge moved through it?

Exercise:

Problem:

To start a car engine, the car battery moves 3.75×10^{21} electrons through the starter motor. How many coulombs of charge were moved?

Solution:

-600 C

Exercise:

Problem:

A certain lightning bolt moves 40.0 C of charge. How many fundamental units of charge $\mid q_e \mid$ is this?

Glossary

electric charge
: a physical property of an object that causes it to be attracted toward or repelled from another charged object; each charged object generates and is influenced by a force called an electromagnetic force

law of conservation of charge
: states that whenever a charge is created, an equal amount of charge with the opposite sign is created simultaneously

electron

	a particle orbiting the nucleus of an atom and carrying the smallest unit of negative charge
proton	a particle in the nucleus of an atom and carrying a positive charge equal in magnitude and opposite in sign to the amount of negative charge carried by an electron

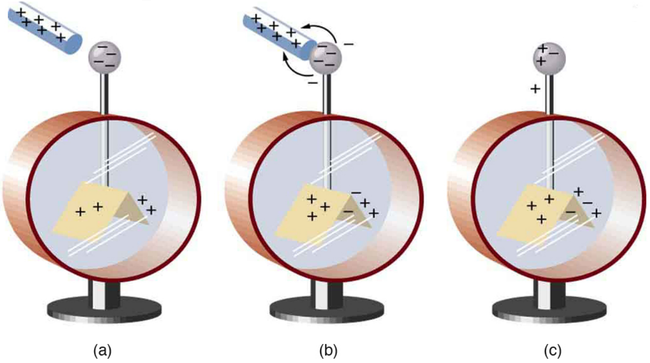
Conductors and Insulators

- Define conductor and insulator, explain the difference, and give examples of each.
- Describe three methods for charging an object.
- Explain what happens to an electric force as you move farther from the source.
- Define polarization.

This power adapter uses metal wires and connectors to conduct electricity from the wall socket to a laptop computer. The conducting wires allow electrons to move freely through the cables, which are shielded by rubber and plastic. These materials act as insulators that don’t allow electric charge to escape outward. (credit: Evan-Amos, Wikimedia Commons)

Some substances, such as metals and salty water, allow charges to move through them with relative ease. Some of the electrons in metals and similar conductors are not bound to individual atoms or sites in the material. These **free electrons** can move through the material much as air moves through loose sand. Any substance that has free electrons and allows charge to move relatively freely through it is called a **conductor**. The moving electrons may collide with fixed atoms and molecules, losing some energy, but they can move in a conductor. Superconductors allow the movement of charge without any loss of energy. Salty water and other similar conducting materials contain free ions that can move through them. An ion is an atom or molecule having a positive or negative (nonzero) total charge. In other words, the total number of electrons is not equal to the total number of protons.

Other substances, such as glass, do not allow charges to move through them. These are called **insulators**. Electrons and ions in insulators are bound in the structure and cannot move easily—as much as 10^{23} times more slowly than in conductors. Pure water and dry table salt are insulators, for example, whereas molten salt and salty water are conductors.


(a) (b) (c)

An electroscope is a favorite instrument in physics demonstrations and student laboratories. It is typically made with gold foil leaves hung from a (conducting) metal stem and is insulated from the room air in a glass-walled container. (a) A positively charged glass rod is brought near the tip of the electroscope, attracting electrons to the top and leaving a net positive charge on the leaves. Like charges in the light flexible gold leaves repel, separating them. (b) When the rod is touched against the ball, electrons are attracted and transferred, reducing the net charge on the glass rod but leaving the electroscope positively charged. (c) The excess charges are evenly distributed in the stem and leaves of the electroscope once the glass rod is removed.

Charging by Contact

[link] shows an electroscope being charged by touching it with a positively charged glass rod. Because the glass rod is an insulator, it must actually touch the electroscope to transfer charge to or from it. (Note that the extra positive charges reside on the surface of the glass rod as a result of rubbing it with silk before starting the experiment.) Since only electrons move in metals, we see that they are attracted to the top of the electroscope. There, some are transferred to the positive rod by touch, leaving the electroscope with a net positive charge.

Electrostatic repulsion in the leaves of the charged electroscope separates them. The electrostatic force has a horizontal component that results in the leaves moving apart as well as a vertical component that is balanced by the gravitational force. Similarly, the electroscope can be negatively charged by contact with a negatively charged object.

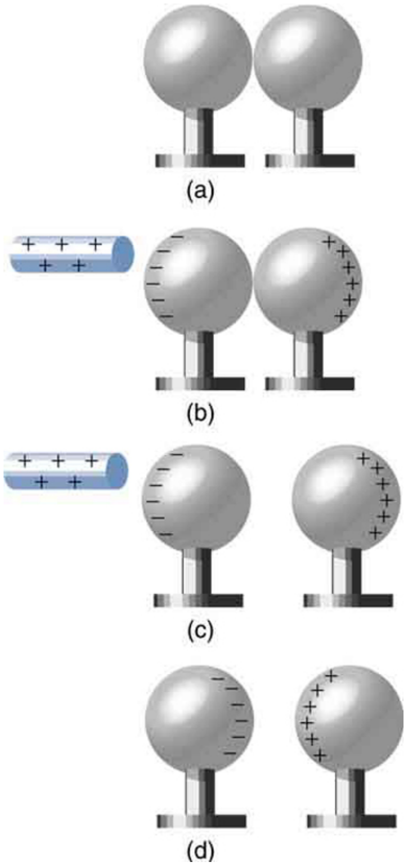
Charging by Induction

It is not necessary to transfer excess charge directly to an object in order to charge it. [link] shows a method of **induction** wherein a charge is created in a nearby object, without direct contact. Here we see two neutral metal spheres in contact with one another but insulated from the rest of the world.

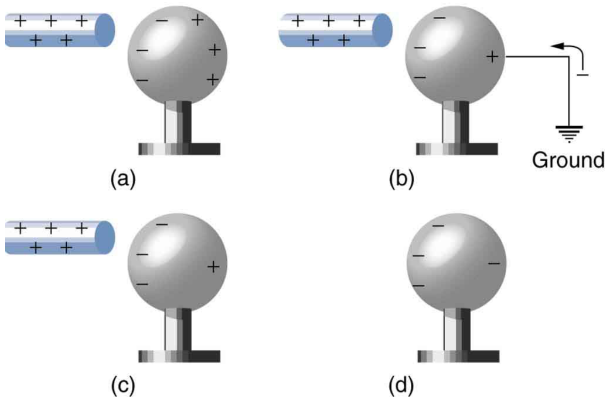
A positively charged rod is brought near one of them, attracting negative charge to that side, leaving the other sphere positively charged.

This is an example of induced **polarization** of neutral objects. Polarization is the separation of charges in an object that remains neutral. If the spheres are now separated (before the rod is pulled away), each sphere will have a net charge. Note that the object closest to the charged rod receives an opposite charge when charged by induction. Note also that no charge is removed from the charged rod, so that this process can be repeated without depleting the supply of excess charge.

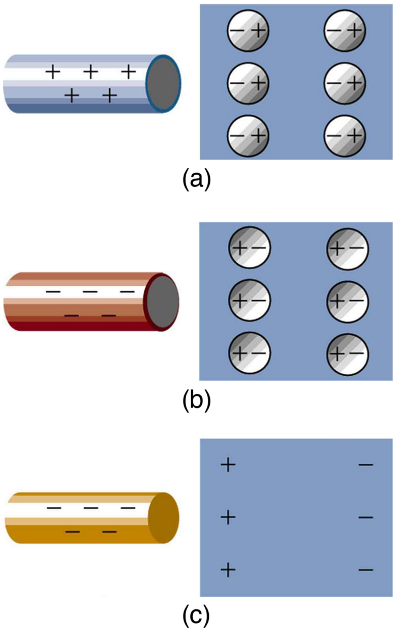
Another method of charging by induction is shown in [link]. The neutral metal sphere is polarized when a charged rod is brought near it. The sphere is then grounded, meaning that a conducting wire is run from the sphere to the ground. Since the earth is large and most ground is a good conductor, it can supply or accept excess charge easily. In this case, electrons are attracted to the sphere through a wire called the ground wire, because it supplies a conducting path to the ground. The ground connection is broken before the charged rod is removed, leaving the sphere with an excess charge opposite to that of the rod. Again, an opposite charge is achieved when charging by induction and the charged rod loses none of its excess charge.



Charging by induction. (a) Two uncharged or neutral metal spheres are in contact with each other but insulated from the rest of the world. (b) A positively charged glass rod is brought near the sphere on the left, attracting negative charge and leaving the other sphere positively charged. (c) The spheres are separated before the rod is removed, thus separating negative and positive charge. (d) The spheres retain net charges after the inducing rod is removed—without ever having been touched by a charged object.



Charging by induction, using a ground connection. (a) A positively charged rod is brought near a neutral metal sphere, polarizing it. (b) The sphere is grounded, allowing electrons to be attracted from the earth’s ample supply. (c) The ground connection is broken. (d) The positive rod is removed, leaving the sphere with an induced negative charge.



Both positive and negative objects attract a neutral object by polarizing its molecules. (a) A positive object brought near a neutral insulator polarizes its molecules. There is a slight shift in the distribution of the electrons orbiting the molecule, with unlike charges being brought nearer and like charges moved away. Since the electrostatic force decreases with distance, there is a net attraction. (b) A negative object produces the opposite polarization, but again attracts the neutral object. (c) The same effect occurs for a conductor; since the unlike charges are closer, there is a net attraction.

Neutral objects can be attracted to any charged object. The pieces of straw attracted to polished amber are neutral, for example. If you run a plastic comb through your hair, the charged comb can pick up neutral pieces of paper. [link] shows how the polarization of atoms and molecules in neutral objects results in their attraction to a charged object.

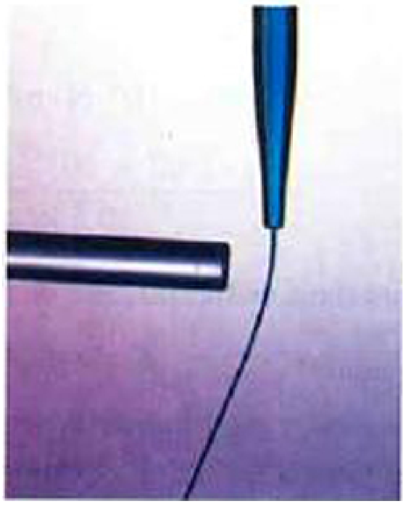
When a charged rod is brought near a neutral substance, an insulator in this case, the distribution of charge in atoms and molecules is shifted slightly. Opposite charge is attracted nearer the external charged rod, while like charge is repelled. Since the electrostatic force decreases with distance, the repulsion of like charges is weaker than the attraction of unlike charges, and so there is a net attraction. Thus a positively charged glass rod attracts neutral pieces of paper, as will a negatively charged rubber rod. Some molecules, like water, are polar molecules. Polar molecules have a natural or inherent separation of charge, although they are neutral overall. Polar molecules are particularly affected by other charged objects and show greater polarization effects than molecules with naturally uniform charge distributions.

Exercise:

Check Your Understanding

Problem:

Can you explain the attraction of water to the charged rod in the figure below?



Solution:

Answer

Water molecules are polarized, giving them slightly positive and slightly negative sides. This makes water even more susceptible to a charged rod’s attraction. As the water flows downward, due to the force of gravity, the charged conductor exerts a net attraction to the opposite charges in the stream of water, pulling it closer.

Note:

PhET Explorations: John Travoltage

Make sparks fly with John Travoltage. Wiggle Johnnie's foot and he picks up charges from the carpet. Bring his hand close to the door knob and get rid of the excess charge.
https://phet.colorado.edu/sims/html/john-travoltage/latest/john-travoltage_en.html

Section Summary

- Polarization is the separation of positive and negative charges in a neutral object.
- A conductor is a substance that allows charge to flow freely through its atomic structure.
- An insulator holds charge within its atomic structure.
- Objects with like charges repel each other, while those with unlike charges attract each other.
- A conducting object is said to be grounded if it is connected to the Earth through a conductor. Grounding allows transfer of charge to and from the earth’s large reservoir.
- Objects can be charged by contact with another charged object and obtain the same sign charge.
- If an object is temporarily grounded, it can be charged by induction, and obtains the opposite sign charge.
- Polarized objects have their positive and negative charges concentrated in different areas, giving them a non-symmetrical charge.
- Polar molecules have an inherent separation of charge.

Conceptual Questions

Exercise:

Problem:

An eccentric inventor attempts to levitate by first placing a large negative charge on himself and then putting a large positive charge on the ceiling of his workshop. Instead, while attempting to place a large negative charge on himself, his clothes fly off. Explain.

Exercise:

Problem:

If you have charged an electroscope by contact with a positively charged object, describe how you could use it to determine the charge of other objects. Specifically, what would the leaves of the electroscope do if other charged objects were brought near its knob?

Exercise:

Problem:

When a glass rod is rubbed with silk, it becomes positive and the silk becomes negative—yet both attract dust. Does the dust have a third type of charge that is attracted to both positive and negative? Explain.

Exercise:

Problem:

Why does a car always attract dust right after it is polished? (Note that car wax and car tires are insulators.)

Exercise:

Problem:

Describe how a positively charged object can be used to give another object a negative charge. What is the name of this process?

Exercise:

Problem:

What is grounding? What effect does it have on a charged conductor? On a charged insulator?

Problems & Exercises

Exercise:

Problem:

Suppose a speck of dust in an electrostatic precipitator has 1.0000×10^{12} protons in it and has a net charge of –5.00 nC (a very large charge for a small speck). How many electrons does it have?

Solution:

1.03×10^{12}

Exercise:

Problem:

An amoeba has 1.00×10^{16} protons and a net charge of 0.300 pC. (a) How many fewer electrons are there than protons? (b) If you paired them up, what fraction of the protons would have no electrons?

Exercise:

Problem:

A 50.0 g ball of copper has a net charge of $2.00\ \mu\text{C}$. What fraction of the copper's electrons has been removed? (Each copper atom has 29 protons, and copper has an atomic mass of 63.5.)

Solution:

9.09×10^{-13}

Exercise:

Problem:

What net charge would you place on a 100 g piece of sulfur if you put an extra electron on 1 in 10^{12} of its atoms? (Sulfur has an atomic mass of 32.1.)

Exercise:

Problem:

How many coulombs of positive charge are there in 4.00 kg of plutonium, given its atomic mass is 244 and that each plutonium atom has 94 protons?

Solution:

$1.48 \times 10^8 \text{ C}$

Glossary

free electron
an electron that is free to move away from its atomic orbit

conductor
a material that allows electrons to move separately from their atomic orbits

insulator
a material that holds electrons securely within their atomic orbits

grounded
when a conductor is connected to the Earth, allowing charge to freely flow to and from Earth’s unlimited reservoir

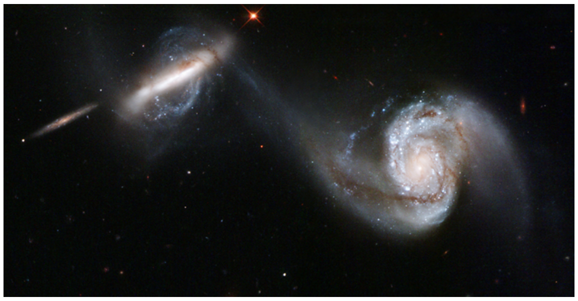
induction
the process by which an electrically charged object brought near a neutral object creates a charge in that object

polarization
slight shifting of positive and negative charges to opposite sides of an atom or molecule

electrostatic repulsion
the phenomenon of two objects with like charges repelling each other

Coulomb’s Law

- State Coulomb’s law in terms of how the electrostatic force changes with the distance between two objects.
- Calculate the electrostatic force between two charged point forces, such as electrons or protons.
- Compare the electrostatic force to the gravitational attraction for a proton and an electron; for a human and the Earth.



This NASA image of Arp 87 shows the result of a strong gravitational attraction between two galaxies. In contrast, at the subatomic level, the electrostatic attraction between two objects, such as an electron and a proton, is far greater than their mutual attraction due to gravity. (credit: NASA/HST)

Through the work of scientists in the late 18th century, the main features of the **electrostatic force**—the existence of two types of charge, the observation that like charges repel, unlike charges attract, and the decrease of force with distance—were eventually refined, and expressed as a mathematical formula. The mathematical formula for the electrostatic force is called **Coulomb’s law** after the French physicist Charles Coulomb (1736–1806), who performed experiments and first proposed a formula to calculate it.

Note:
Coulomb’s Law
Equation:

$$F = k\frac{|q_1 q_2|}{r^2}.$$

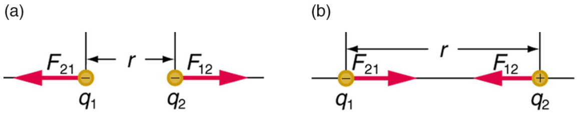
Coulomb's law calculates the magnitude of the force F between two point charges, q_1 and q_2, separated by a distance r. In SI units, the constant k is equal to

Equation:

$$k = 8.988 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2} \approx 8.99 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2}.$$

The electrostatic force is a vector quantity and is expressed in units of newtons. The force is understood to be along the line joining the two charges. (See [link].)

Although the formula for Coulomb's law is simple, it was no mean task to prove it. The experiments Coulomb did, with the primitive equipment then available, were difficult. Modern experiments have verified Coulomb's law to great precision. For example, it has been shown that the force is inversely proportional to distance between two objects squared $\left(F \propto 1/r^2\right)$ to an accuracy of 1 part in 10^{16}. No exceptions have ever been found, even at the small distances within the atom.



The magnitude of the electrostatic force F between point charges q_1 and q_2 separated by a distance r is given by Coulomb's law. Note that Newton's third law (every force exerted creates an equal and opposite force) applies as usual—the force on q_1 is equal in magnitude and opposite in direction to the force it exerts on q_2. (a) Like charges. (b) Unlike charges.

Example:

How Strong is the Coulomb Force Relative to the Gravitational Force?

Compare the electrostatic force between an electron and proton separated by 0.530×10^{-10} m with the gravitational force between them. This distance is their average separation in a hydrogen atom.

Strategy

To compare the two forces, we first compute the electrostatic force using Coulomb's law, $F = k\frac{|q_1 q_2|}{r^2}$. We then calculate the gravitational force using Newton's universal law of gravitation. Finally, we take a ratio to see how the forces compare in magnitude.

Solution

Entering the given and known information about the charges and separation of the electron and proton into the expression of Coulomb's law yields

Equation:

$$F = k\frac{|q_1 q_2|}{r^2}$$

Equation:

$$= \left(8.99 \times 10^9\ \text{N} \cdot \text{m}^2/\text{C}^2\right) \times \frac{(1.60 \times 10^{-19}\ \text{C})(1.60 \times 10^{-19}\ \text{C})}{(0.530 \times 10^{-10}\ \text{m})^2}$$

Thus the Coulomb force is

Equation:

$$F = 8.19 \times 10^{-8}\ \text{N}.$$

The charges are opposite in sign, so this is an attractive force. This is a very large force for an electron—it would cause an acceleration of $8.99 \times 10^{22}\ \text{m/s}^2$(verification is left as an end-of-section problem).The gravitational force is given by Newton's law of gravitation as:

Equation:

$$F_G = G\frac{mM}{r^2},$$

where $G = 6.67 \times 10^{-11}\ \text{N} \cdot \text{m}^2/\text{kg}^2$. Here m and M represent the electron and proton masses, which can be found in the appendices. Entering values for the knowns yields

Equation:

$$F_G = (6.67 \times 10^{-11}\ \text{N} \cdot \text{m}^2/\text{kg}^2) \times \frac{(9.11 \times 10^{-31}\ \text{kg})(1.67 \times 10^{-27}\ \text{kg})}{(0.530 \times 10^{-10}\ \text{m})^2} = 3.61 \times 10^{-47}\ \text{N}$$

This is also an attractive force, although it is traditionally shown as positive since gravitational force is always attractive. The ratio of the magnitude of the electrostatic force to gravitational force in this case is, thus,

Equation:

$$\frac{F}{F_G} = 2.27 \times 10^{39}.$$

Discussion

This is a remarkably large ratio! Note that this will be the ratio of electrostatic force to gravitational force for an electron and a proton at any distance (taking the ratio before entering numerical values shows that the distance cancels). This ratio gives some indication of just how much larger the Coulomb force is than the gravitational force between two of the most common particles in nature.

As the example implies, gravitational force is completely negligible on a small scale, where the interactions of individual charged particles are important. On a large scale, such as between the Earth and a person, the reverse is true. Most objects are nearly electrically neutral, and so attractive and repulsive **Coulomb forces** nearly cancel. Gravitational force on a large scale dominates interactions between large objects because it is always attractive, while Coulomb forces tend to cancel.

Section Summary

- Frenchman Charles Coulomb was the first to publish the mathematical equation that describes the electrostatic force between two objects.
- Coulomb’s law gives the magnitude of the force between point charges. It is
  Equation:

  $$F = k\frac{|q_1 q_2|}{r^2},$$

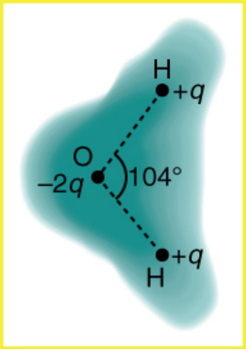
  where q_1 and q_2 are two point charges separated by a distance r, and $k \approx 8.99 \times 10^9\ \mathrm{N \cdot m^2/C^2}$
- This Coulomb force is extremely basic, since most charges are due to point-like particles. It is responsible for all electrostatic effects and underlies most macroscopic forces.
- The Coulomb force is extraordinarily strong compared with the gravitational force, another basic force—but unlike gravitational force it can cancel, since it can be either attractive or repulsive.
- The electrostatic force between two subatomic particles is far greater than the gravitational force between the same two particles.

Conceptual Questions

Exercise:

Problem:

[link] shows the charge distribution in a water molecule, which is called a polar molecule because it has an inherent separation of charge. Given water’s polar character, explain what effect humidity has on removing excess charge from objects.



Schematic representation of the outer electron cloud of a neutral water molecule. The electrons spend more time near the oxygen than the hydrogens, giving a permanent charge separation as shown. Water is thus a *polar molecule*. It is more easily affected by electrostatic forces than molecules with uniform charge distributions.

Exercise:

Problem:

Using [link], explain, in terms of Coulomb’s law, why a polar molecule (such as in [link]) is attracted by both positive and negative charges.

Exercise:

Problem:

Given the polar character of water molecules, explain how ions in the air form nucleation centers for rain droplets.

Problems & Exercises

Exercise:

Problem:

What is the repulsive force between two pith balls that are 8.00 cm apart and have equal charges of – 30.0 nC?

Exercise:

Problem:

(a) How strong is the attractive force between a glass rod with a $0.700\ \mu\text{C}$ charge and a silk cloth with a –$0.600\ \mu\text{C}$ charge, which are 12.0 cm apart, using the approximation that they act like point charges? (b) Discuss how the answer to this problem might be affected if the charges are distributed over some area and do not act like point charges.

Solution:

(a) 0.263 N

(b) If the charges are distributed over some area, there will be a concentration of charge along the side closest to the oppositely charged object. This effect will increase the net force.

Exercise:

Problem:

Two point charges exert a 5.00 N force on each other. What will the force become if the distance between them is increased by a factor of three?

Exercise:

Problem:

Two point charges are brought closer together, increasing the force between them by a factor of 25. By what factor was their separation decreased?

Solution:

The separation decreased by a factor of 5.

Exercise:

Problem:

How far apart must two point charges of 75.0 nC (typical of static electricity) be to have a force of 1.00 N between them?

Exercise:

Problem:

If two equal charges each of 1 C each are separated in air by a distance of 1 km, what is the magnitude of the force acting between them? You will see that even at a distance as large as 1 km, the repulsive force is substantial because 1 C is a very significant amount of charge.

Exercise:

Problem:

A test charge of $+2\ \mu\text{C}$ is placed halfway between a charge of $+6\ \mu\text{C}$ and another of $+4\ \mu\text{C}$ separated by 10 cm. (a) What is the magnitude of the force on the test charge? (b) What is the direction of this force (away from or toward the $+6\ \mu\text{C}$ charge)?

Exercise:

Problem:

Bare free charges do not remain stationary when close together. To illustrate this, calculate the acceleration of two isolated protons separated by 2.00 nm (a typical distance between gas atoms). Explicitly show how you follow the steps in the Problem-Solving Strategy for electrostatics.

Solution:

$$\begin{aligned} F &= k\frac{|q_1q_2|}{r^2} = ma \Rightarrow a = \frac{kq^2}{mr^2} \\ &= \frac{\left(9.00\times10^{9}\ \text{N}\cdot\text{m}^2/\text{C}^2\right)\left(1.60\times10^{-19}\ \text{m}\right)^2}{\left(1.67\times10^{-27}\ \text{kg}\right)\left(2.00\times10^{-9}\ \text{m}\right)^2} \\ &= 3.45\times10^{16}\ \text{m/s}^2 \end{aligned}$$

Exercise:

Problem:

(a) By what factor must you change the distance between two point charges to change the force between them by a factor of 10? (b) Explain how the distance can either increase or decrease by this factor and still cause a factor of 10 change in the force.

Solution:

(a) 3.2

(b) If the distance increases by 3.2, then the force will decrease by a factor of 10 ; if the distance decreases by 3.2, then the force will increase by a factor of 10. Either way, the force changes by a factor of 10.

Exercise:

Problem:

Suppose you have a total charge q_{tot} that you can split in any manner. Once split, the separation distance is fixed. How do you split the charge to achieve the greatest force?

Exercise:

Problem:

(a) Common transparent tape becomes charged when pulled from a dispenser. If one piece is placed above another, the repulsive force can be great enough to support the top piece’s weight. Assuming equal point charges (only an approximation), calculate the magnitude of the charge if electrostatic force is great enough to support the weight of a 10.0 mg piece of tape held 1.00 cm above another. (b) Discuss whether the magnitude of this charge is consistent with what is typical of static electricity.

Solution:

(a) 1.04×10^{-9} C

(b) This charge is approximately 1 nC, which is consistent with the magnitude of charge typical for static electricity

Exercise:

Problem:

(a) Find the ratio of the electrostatic to gravitational force between two electrons. (b) What is this ratio for two protons? (c) Why is the ratio different for electrons and protons?

Exercise:

Problem:

At what distance is the electrostatic force between two protons equal to the weight of one proton?

Exercise:

Problem:

A certain five cent coin contains 5.00 g of nickel. What fraction of the nickel atoms' electrons, removed and placed 1.00 m above it, would support the weight of this coin? The atomic mass of nickel is 58.7, and each nickel atom contains 28 electrons and 28 protons.

Solution:

1.02×10^{-11}

Exercise:

Problem:

(a) Two point charges totaling $8.00\ \mu\text{C}$ exert a repulsive force of 0.150 N on one another when separated by 0.500 m. What is the charge on each? (b) What is the charge on each if the force is attractive?

Exercise:

Problem:

Point charges of $5.00\ \mu\text{C}$ and $-3.00\ \mu\text{C}$ are placed 0.250 m apart. (a) Where can a third charge be placed so that the net force on it is zero? (b) What if both charges are positive?

Solution:

a. 0.859 m beyond negative charge on line connecting two charges
b. 0.109 m from lesser charge on line connecting two charges

Exercise:

Problem:

Two point charges q_1 and q_2 are $3.00\ \text{m}$ apart, and their total charge is $20\ \mu\text{C}$. (a) If the force of repulsion between them is 0.075N, what are magnitudes of the two charges? (b) If one charge attracts the other with a force of 0.525N, what are the magnitudes of the two charges? Note that you may need to solve a quadratic equation to reach your answer.

Glossary

Coulomb's law	the mathematical equation calculating the electrostatic force vector between two charged particles
Coulomb force	another term for the electrostatic force

electrostatic force

the amount and direction of attraction or repulsion between two charged bodies

Electric Field: Concept of a Field Revisited

- Describe a force field and calculate the strength of an electric field due to a point charge.
- Calculate the force exerted on a test charge by an electric field.
- Explain the relationship between electrical force (F) on a test charge and electrical field strength (E).

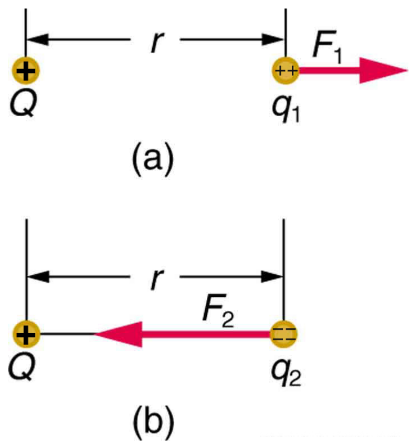
Contact forces, such as between a baseball and a bat, are explained on the small scale by the interaction of the charges in atoms and molecules in close proximity. They interact through forces that include the **Coulomb force**. Action at a distance is a force between objects that are not close enough for their atoms to “touch.” That is, they are separated by more than a few atomic diameters.

For example, a charged rubber comb attracts neutral bits of paper from a distance via the Coulomb force. It is very useful to think of an object being surrounded in space by a **force field**. The force field carries the force to another object (called a test object) some distance away.

Concept of a Field

A field is a way of conceptualizing and mapping the force that surrounds any object and acts on another object at a distance without apparent physical connection. For example, the gravitational field surrounding the earth (and all other masses) represents the gravitational force that would be experienced if another mass were placed at a given point within the field.

In the same way, the Coulomb force field surrounding any charge extends throughout space. Using Coulomb’s law, $F=k|q_1q_2|/r^2$, its magnitude is given by the equation $F=k|\ qQ\ |/r^2$, for a **point charge** (a particle having a charge Q) acting on a **test charge** q at a distance r (see [link]). Both the magnitude and direction of the Coulomb force field depend on Q and the test charge q.



The Coulomb force field due to a positive charge Q is shown acting on two different charges. Both charges are the same distance from Q. (a) Since q_1 is positive, the force F_1 acting on it is repulsive. (b) The charge q_2 is negative and greater in magnitude than q_1, and so the force F_2 acting on it is attractive and stronger than F_1. The Coulomb force field is thus not unique at any point in space, because it depends on the test charges q_1 and q_2

as well as the
charge Q.

To simplify things, we would prefer to have a field that depends only on Q and not on the test charge q. The electric field is defined in such a manner that it represents only the charge creating it and is unique at every point in space. Specifically, the electric field E is defined to be the ratio of the Coulomb force to the test charge:

Equation:

$$\mathbf{E} = \frac{\mathbf{F}}{q},$$

where $\mathbf{F}$ is the electrostatic force (or Coulomb force) exerted on a positive test charge q. It is understood that $\mathbf{E}$ is in the same direction as $\mathbf{F}$. It is also assumed that q is so small that it does not alter the charge distribution creating the electric field. The units of electric field are newtons per coulomb (N/C). If the electric field is known, then the electrostatic force on any charge q is simply obtained by multiplying charge times electric field, or $\mathbf{F} = q\mathbf{E}$. Consider the electric field due to a point charge Q. According to Coulomb's law, the force it exerts on a test charge q is $F = k|\,qQ\,|/r^2$. Thus the magnitude of the electric field, E, for a point charge is

Equation:

$$E = \left|\frac{F}{q}\right| = k\left|\frac{\text{qQ}}{qr^2}\right| = k\frac{|Q|}{r^2}.$$

Since the test charge cancels, we see that

Equation:

$$E = k\frac{|Q|}{r^2}.$$

The electric field is thus seen to depend only on the charge Q and the distance r; it is completely independent of the test charge q.

Example:

Calculating the Electric Field of a Point Charge

Calculate the strength and direction of the electric field E due to a point charge of 2.00 nC (nano-Coulombs) at a distance of 5.00 mm from the charge.

Strategy

We can find the electric field created by a point charge by using the equation $E = \mathrm{kQ}/r^2$.

Solution

Here $Q = 2.00 \times 10^{-9}$ C and $r = 5.00 \times 10^{-3}$ m. Entering those values into the above equation gives

Equation:

$$\begin{aligned} E &= k\frac{Q}{r^2} \\ &= (8.99 \times 10^9\ \mathrm{N \cdot m^2/C^2}) \times \frac{(2.00\times10^{-9}\ \mathrm{C})}{(5.00\times10^{-3}\ \mathrm{m})^2} \\ &= 7.19 \times 10^5\ \mathrm{N/C}. \end{aligned}$$

Discussion

This **electric field strength** is the same at any point 5.00 mm away from the charge Q that creates the field. It is positive, meaning that it has a direction pointing away from the charge Q.

Example:

Calculating the Force Exerted on a Point Charge by an Electric Field

What force does the electric field found in the previous example exert on a point charge of –0.250 μC?

Strategy

Since we know the electric field strength and the charge in the field, the force on that charge can be calculated using the definition of electric field

$\mathbf{E} = \mathbf{F}/q$ rearranged to $\mathbf{F} = q\mathbf{E}$.

Solution

The magnitude of the force on a charge $q = -0.250\ \mu\text{C}$ exerted by a field of strength $E = 7.20 \times 10^5$ N/C is thus,

Equation:

$$\begin{aligned} F &= -qE \\ &= (0.250 \times 10^{-6}\ \text{C})(7.20 \times 10^5\ \text{N/C}) \\ &= 0.180\ \text{N}. \end{aligned}$$

Because q is negative, the force is directed opposite to the direction of the field.

Discussion

The force is attractive, as expected for unlike charges. (The field was created by a positive charge and here acts on a negative charge.) The charges in this example are typical of common static electricity, and the modest attractive force obtained is similar to forces experienced in static cling and similar situations.

Note:

PhET Explorations: Electric Field of Dreams

Play ball! Add charges to the Field of Dreams and see how they react to the electric field. Turn on a background electric field and adjust the direction and magnitude.

https://archive.cnx.org/specials/ca9a78b4-06a7-11e6-b638-3bb71d1f0b42/electric-field-of-dreams/#sim-electric-field-of-dreams

Section Summary

- The electrostatic force field surrounding a charged object extends out into space in all directions.
- The electrostatic force exerted by a point charge on a test charge at a distance r depends on the charge of both charges, as well as the

distance between the two.

- The electric field $\mathbf{E}$ is defined to be
  Equation:

  $$\mathbf{E} = \frac{\mathbf{F}}{q,}$$

  where $\mathbf{F}$ is the Coulomb or electrostatic force exerted on a small positive test charge q. $\mathbf{E}$ has units of N/C.
- The magnitude of the electric field $\mathbf{E}$ created by a point charge Q is
  Equation:

  $$\mathbf{E} = k\frac{|Q|}{r^2}.$$

  where r is the distance from Q. The electric field $\mathbf{E}$ is a vector and fields due to multiple charges add like vectors.

Conceptual Questions

Exercise:

Problem:

Why must the test charge q in the definition of the electric field be vanishingly small?

Exercise:

Problem:

Are the direction and magnitude of the Coulomb force unique at a given point in space? What about the electric field?

Problem Exercises

Exercise:

Problem:

What is the magnitude and direction of an electric field that exerts a 2.00×10^{-5} N upward force on a –$1.75\ \mu$C charge?

Exercise:

Problem:

What is the magnitude and direction of the force exerted on a $3.50\ \mu$C charge by a 250 N/C electric field that points due east?

Solution:

8.75×10^{-4} N

Exercise:

Problem:

Calculate the magnitude of the electric field 2.00 m from a point charge of 5.00 mC (such as found on the terminal of a Van de Graaff).

Exercise:

Problem:

(a) What magnitude point charge creates a 10,000 N/C electric field at a distance of 0.250 m? (b) How large is the field at 10.0 m?

Solution:

(a) 6.94×10^{-8} C

(b) $6.25\ \text{N/C}$

Exercise:

Problem:

Calculate the initial (from rest) acceleration of a proton in a $5.00 \times 10^6 \text{ N/C}$ electric field (such as created by a research Van de Graaff). Explicitly show how you follow the steps in the Problem-Solving Strategy for electrostatics.

Exercise:

Problem:

(a) Find the magnitude and direction of an electric field that exerts a $4.80 \times 10^{-17} \text{ N}$ westward force on an electron. (b) What magnitude and direction force does this field exert on a proton?

Solution:

(a) 300 N/C (east)

(b) $4.80 \times 10^{-17} \text{ N (east)}$

Glossary

field
: a map of the amount and direction of a force acting on other objects, extending out into space

point charge
: A charged particle, designated Q, generating an electric field

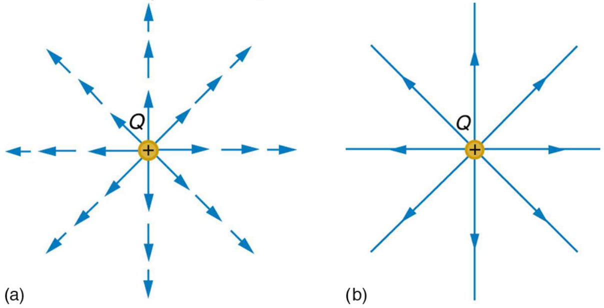
test charge
: A particle (designated q) with either a positive or negative charge set down within an electric field generated by a point charge

Electric Field Lines: Multiple Charges

- Calculate the total force (magnitude and direction) exerted on a test charge from more than one charge
- Describe an electric field diagram of a positive point charge; of a negative point charge with twice the magnitude of positive charge
- Draw the electric field lines between two points of the same charge; between two points of opposite charge.

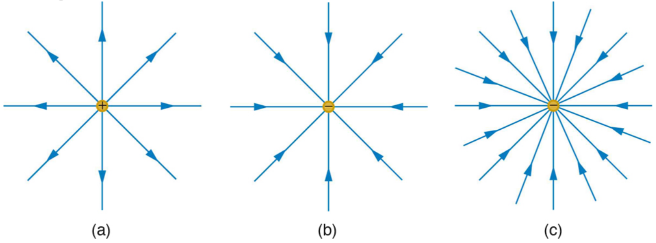
Drawings using lines to represent **electric fields** around charged objects are very useful in visualizing field strength and direction. Since the electric field has both magnitude and direction, it is a vector. Like all **vectors**, the electric field can be represented by an arrow that has length proportional to its magnitude and that points in the correct direction. (We have used arrows extensively to represent force vectors, for example.)

[link] shows two pictorial representations of the same electric field created by a positive point charge Q. [link] (b) shows the standard representation using continuous lines. [link] (a) shows numerous individual arrows with each arrow representing the force on a test charge q. Field lines are essentially a map of infinitesimal force vectors.



Two equivalent representations of the electric field due to a positive charge Q. (a) Arrows representing the electric field's magnitude and direction. (b) In the standard representation, the arrows are replaced by continuous field lines having the same direction at any point as the electric field. The closeness of the lines is directly related to the strength of the electric field. A test charge placed anywhere will feel a force in the direction of the field line; this force will have a strength proportional to the density of the lines (being greater near the charge, for example).

Note that the electric field is defined for a positive test charge q, so that the field lines point away from a positive charge and toward a negative charge. (See [link].) The electric field strength is exactly proportional to the number of field lines per unit area, since the magnitude of the electric field for a point charge is $E = k|Q|/r^2$ and area is proportional to r^2. This pictorial representation, in which field lines represent the direction and their closeness (that is, their areal density or the number of lines crossing a unit area) represents strength, is used for all fields: electrostatic, gravitational, magnetic, and others.


(a)
(b)
(c)

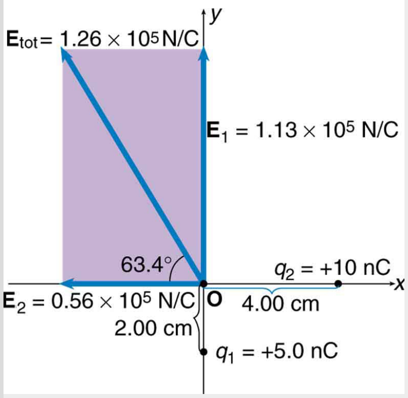
The electric field surrounding three different point charges. (a) A positive charge. (b) A negative charge of equal magnitude. (c) A larger negative charge.

In many situations, there are multiple charges. The total electric field created by multiple charges is the vector sum of the individual fields created by each charge. The following example shows how to add electric field vectors.

Example:

Adding Electric Fields

Find the magnitude and direction of the total electric field due to the two point charges, q_1 and q_2, at the origin of the coordinate system as shown in [link].



y
$E_{tot} = 1.26 \times 10^5$ N/C
$E_1 = 1.13 \times 10^5$ N/C
63.4°
$q_2 = +10$ nC
x
$E_2 = 0.56 \times 10^5$ N/C
O
4.00 cm
2.00 cm
$q_1 = +5.0$ nC

The electric fields $\mathbf{E}_1$ and $\mathbf{E}_2$ at the origin O add to $\mathbf{E}_{\text{tot}}$.

Strategy

Since the electric field is a vector (having magnitude and direction), we add electric fields with the same vector techniques used for other types of vectors. We first must find the electric field due to each charge at the point of interest, which is the origin of the coordinate system (O) in this instance. We pretend that there is a positive test charge, q, at point O, which allows us to determine the direction of the fields $\mathbf{E}_1$ and $\mathbf{E}_2$. Once those fields are found, the total field can be determined using **vector addition**.

Solution

The electric field strength at the origin due to q_1 is labeled E_1 and is calculated:

Equation:

$$E_1 = k\frac{q_1}{r_1^2} = \left(8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2\right)\frac{(5.00\times 10^{-9}\text{ C})}{(2.00\times 10^{-2}\text{ m})^2}$$
$$E_1 = 1.124 \times 10^5 \text{ N/C}.$$

Similarly, E_2 is

Equation:

$$E_2 = k\frac{q_2}{r_2^2} = \left(8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2\right)\frac{(10.0\times 10^{-9}\text{ C})}{(4.00\times 10^{-2}\text{ m})^2}$$
$$E_2 = 0.5619 \times 10^5 \text{ N/C}.$$

Four digits have been retained in this solution to illustrate that E_1 is exactly twice the magnitude of E_2. Now arrows are drawn to represent the magnitudes and directions of $\mathbf{E}_1$ and $\mathbf{E}_2$. (See [link].) The direction of the electric field is that of the force on a positive charge so both arrows point directly away from the positive charges that create them. The arrow for $\mathbf{E}_1$ is exactly twice the length of that for $\mathbf{E}_2$. The arrows form a right triangle in this case and can be added using the Pythagorean theorem. The magnitude of the total field E_{tot} is

Equation:

$$\begin{aligned} E_{\text{tot}} &= (E_1^2 + E_2^2)^{1/2} \\ &= \{(1.124 \times 10^5 \text{ N/C})^2 + (0.5619 \times 10^5 \text{ N/C})^2\}^{1/2} \\ &= 1.26 \times 10^5 \text{ N/C}. \end{aligned}$$

The direction is

Equation:

$$\begin{array}{rcl}\theta & = & \tan^{-1}\left(\frac{E_1}{E_2}\right) \\ & = & \tan^{-1}\left(\frac{1.124\times10^5\text{ N/C}}{0.5619\times10^5\text{ N/C}}\right) \\ & = & 63.4^{\text{o}},\end{array}$$

or 63.4^{o} above the *x*-axis.

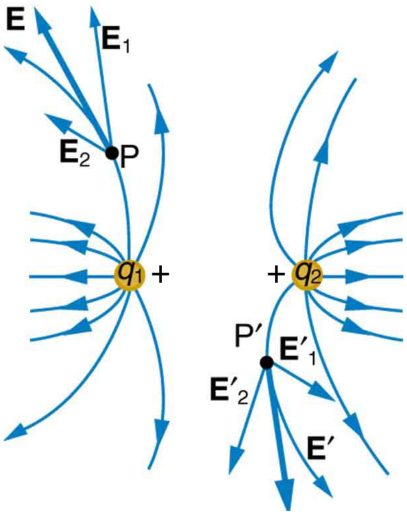
Discussion

In cases where the electric field vectors to be added are not perpendicular, vector components or graphical techniques can be used. The total electric field found in this example is the total electric field at only one point in space. To find the total electric field due to these two charges over an entire region, the same technique must be repeated for each point in the region. This impossibly lengthy task (there are an infinite number of points in space) can be avoided by calculating the total field at representative points and using some of the unifying features noted next.

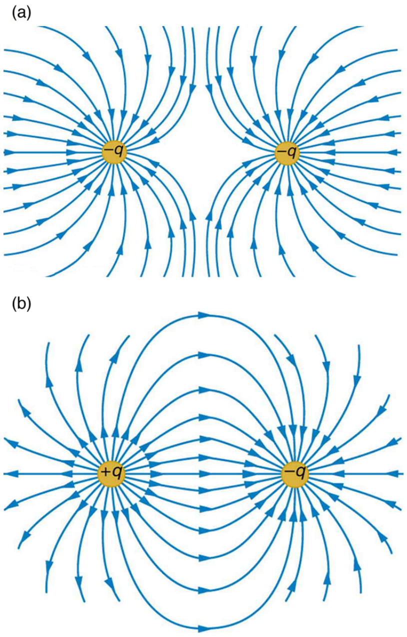
[link] shows how the electric field from two point charges can be drawn by finding the total field at representative points and drawing electric field lines consistent with those points. While the electric fields from multiple charges are more complex than those of single charges, some simple features are easily noticed.

For example, the field is weaker between like charges, as shown by the lines being farther apart in that region. (This is because the fields from each charge exert opposing forces on any charge placed between them.) (See [link] and [link](a).) Furthermore, at a great distance from two like charges, the field becomes identical to the field from a single, larger charge.

[link](b) shows the electric field of two unlike charges. The field is stronger between the charges. In that region, the fields from each charge are in the same direction, and so their strengths add. The field of two unlike charges is weak at large distances, because the fields of the individual charges are in opposite directions and so their strengths subtract. At very large distances, the field of two unlike charges looks like that of a smaller single charge.



Two positive point charges q_1 and q_2 produce the resultant electric field shown. The field is calculated at representative points and then smooth field lines drawn following the rules outlined in the text.



(a) Two negative charges produce the fields shown. It is very similar to the field produced by two positive charges, except that the directions are reversed. The field is clearly weaker between the charges. The individual forces on a test charge in that region are in opposite directions. (b) Two opposite charges produce the field shown, which is stronger in the region between the charges.

We use electric field lines to visualize and analyze electric fields (the lines are a pictorial tool, not a physical entity in themselves). The properties of electric field lines for any charge distribution can be summarized as follows:

1. Field lines must begin on positive charges and terminate on negative charges, or at infinity in the hypothetical case of isolated charges.
2. The number of field lines leaving a positive charge or entering a negative charge is proportional to the magnitude of the charge.
3. The strength of the field is proportional to the closeness of the field lines—more precisely, it is proportional to the number of lines per unit area perpendicular to the lines.
4. The direction of the electric field is tangent to the field line at any point in space.
5. Field lines can never cross.

The last property means that the field is unique at any point. The field line represents the direction of the field; so if they crossed, the field would have two directions at that location (an impossibility if the field is unique).

Note:
PhET Explorations: Charges and Fields
Move point charges around on the playing field and then view the electric field, voltages, equipotential lines, and more. It's colorful, it's dynamic, it's free.

Click here for the simulation.

Section Summary

- Drawings of electric field lines are useful visual tools. The properties of electric field lines for any charge distribution are that:
- Field lines must begin on positive charges and terminate on negative charges, or at infinity in the hypothetical case of isolated charges.
- The number of field lines leaving a positive charge or entering a negative charge is proportional to the magnitude of the charge.
- The strength of the field is proportional to the closeness of the field lines—more precisely, it is proportional to the number of lines per unit area perpendicular to the lines.
- The direction of the electric field is tangent to the field line at any point in space.
- Field lines can never cross.

Conceptual Questions

Exercise:

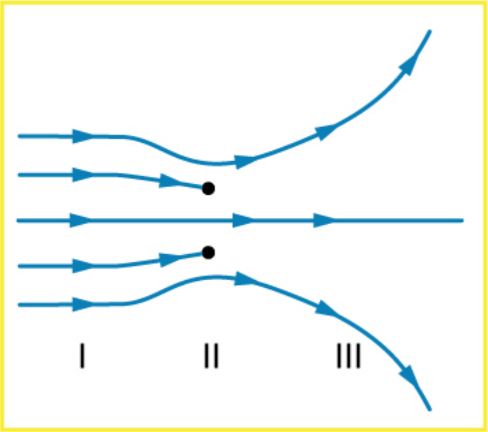
Problem:

Compare and contrast the Coulomb force field and the electric field. To do this, make a list of five properties for the Coulomb force field analogous to the five properties listed for electric field lines. Compare each item in your list of Coulomb force field properties with those of the electric field—are they the same or different? (For example, electric field lines cannot cross. Is the same true for Coulomb field lines?)

Exercise:

Problem:

[link] shows an electric field extending over three regions, labeled I, II, and III. Answer the following questions. (a) Are there any isolated charges? If so, in what region and what are their signs? (b) Where is the field strongest? (c) Where is it weakest? (d) Where is the field the most uniform?



Problem Exercises

Exercise:

Problem:

(a) Sketch the electric field lines near a point charge $+q$. (b) Do the same for a point charge $-3.00q$.

Exercise:

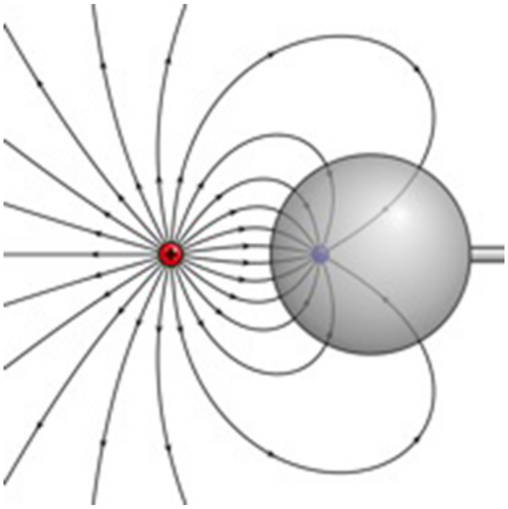
Problem:

Sketch the electric field lines a long distance from the charge distributions shown in [link] (a) and (b)

Exercise:

Problem:

[link] shows the electric field lines near two charges q_1 and q_2. What is the ratio of their magnitudes? (b) Sketch the electric field lines a long distance from the charges shown in the figure.



The electric field near two charges.

Exercise:

Problem:

Sketch the electric field lines in the vicinity of two opposite charges, where the negative charge is three times greater in magnitude than the positive. (See [link] for a similar situation).

Glossary

electric field	a three-dimensional map of the electric force extended out into space from a point charge
electric field lines	a series of lines drawn from a point charge representing the magnitude and direction of force exerted by that charge
vector	a quantity with both magnitude and direction

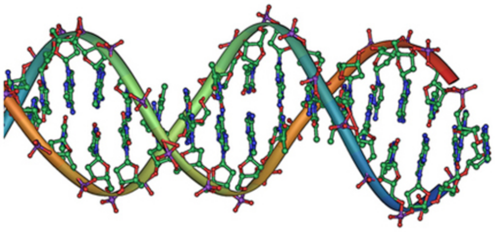
vector addition

mathematical combination of two or more vectors, including their magnitudes, directions, and positions

Electric Forces in Biology

- Describe how a water molecule is polar.
- Explain electrostatic screening by a water molecule within a living cell.

Classical electrostatics has an important role to play in modern molecular biology. Large molecules such as proteins, nucleic acids, and so on—so important to life—are usually electrically charged. DNA itself is highly charged; it is the electrostatic force that not only holds the molecule together but gives the molecule structure and strength. [link] is a schematic of the DNA double helix.



DNA is a highly charged molecule. The DNA double helix shows the two coiled strands each containing a row of nitrogenous bases, which “code” the genetic information needed by a living organism. The strands are connected by bonds between pairs of bases. While pairing combinations between certain bases are fixed (C-G and A-T), the sequence of nucleotides in the strand varies. (credit: Jerome Walker)

The four nucleotide bases are given the symbols A (adenine), C (cytosine), G (guanine), and T (thymine). The order of the four bases varies in each strand, but the pairing between bases is always the same. C and G are always paired and A and T are always paired, which helps to preserve the order of bases in cell division (mitosis) so as to pass on the correct genetic information. Since the Coulomb force drops with distance ($F \propto 1/r^2$), the distances between the base pairs must be small enough that the electrostatic force is sufficient to hold them together.

DNA is a highly charged molecule, with about $2q_{\text{e}}$ (fundamental charge) per 0.3×10^{-9} m. The distance separating the two strands that make up the DNA structure is about 1 nm, while the distance separating the individual atoms within each base is about 0.3 nm.

One might wonder why electrostatic forces do not play a larger role in biology than they do if we have so many charged molecules. The reason is that the electrostatic force is “diluted” due to **screening** between molecules. This is due to the presence of other charges in the cell.

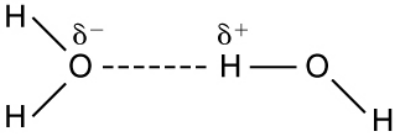
Polarity of Water Molecules

The best example of this charge screening is the water molecule, represented as H_2O. Water is a strongly **polar molecule**. Its 10 electrons (8 from the oxygen atom and 2 from the two hydrogen atoms) tend to remain closer to the oxygen nucleus than the hydrogen nuclei. This creates two centers of equal and opposite charges—what is called a **dipole**, as illustrated in [link]. The magnitude of the dipole is called the dipole moment.

These two centers of charge will terminate some of the electric field lines coming from a free charge, as on a DNA molecule. This results in a reduction in the strength of the **Coulomb interaction**. One might say that screening makes the Coulomb force a short range force rather than long range.

Other ions of importance in biology that can reduce or screen Coulomb interactions are Na^+, and K^+, and Cl^-. These ions are located both inside and outside of living cells. The movement of these ions through cell membranes is crucial to the motion of nerve impulses through nerve axons.

Recent studies of electrostatics in biology seem to show that electric fields in cells can be extended over larger distances, in spite of screening, by “microtubules” within the cell. These microtubules are hollow tubes composed of proteins that guide the movement of chromosomes when cells divide, the motion of other organisms within the cell, and provide mechanisms for motion of some cells (as motors).



This schematic shows water (H_2O) as a polar molecule. Unequal sharing of electrons between the oxygen (O) and hydrogen (H) atoms leads to a net separation of positive and negative charge—forming a dipole. The symbols δ^- and δ^+ indicate that the oxygen side of the H_2O molecule tends to be more negative, while the hydrogen ends tend

to be more positive.
This leads to an
attraction of
opposite charges
between molecules.

Section Summary

- Many molecules in living organisms, such as DNA, carry a charge.
- An uneven distribution of the positive and negative charges within a polar molecule produces a dipole.
- The effect of a Coulomb field generated by a charged object may be reduced or blocked by other nearby charged objects.
- Biological systems contain water, and because water molecules are polar, they have a strong effect on other molecules in living systems.

Conceptual Question

Exercise:

Problem:

A cell membrane is a thin layer enveloping a cell. The thickness of the membrane is much less than the size of the cell. In a static situation the membrane has a charge distribution of $-2.5 \times 10^{-6} \text{C/m}^2$ on its inner surface and $+2.5 \times 10^{-6}\ \text{C/m}^2$ on its outer surface. Draw a diagram of the cell and the surrounding cell membrane. Include on this diagram the charge distribution and the corresponding electric field. Is there any electric field inside the cell? Is there any electric field outside the cell?

Glossary

dipole

a molecule’s lack of symmetrical charge distribution, causing one side to be more positive and another to be more negative

polar molecule
a molecule with an asymmetrical distribution of positive and negative charge

screening
the dilution or blocking of an electrostatic force on a charged object by the presence of other charges nearby

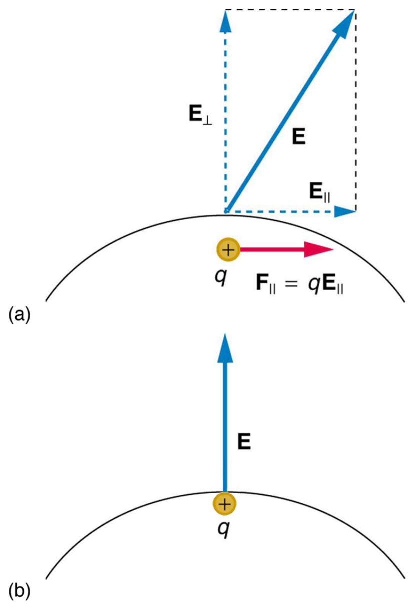
Coulomb interaction
the interaction between two charged particles generated by the Coulomb forces they exert on one another

Conductors and Electric Fields in Static Equilibrium

- List the three properties of a conductor in electrostatic equilibrium.
- Explain the effect of an electric field on free charges in a conductor.
- Explain why no electric field may exist inside a conductor.
- Describe the electric field surrounding Earth.
- Explain what happens to an electric field applied to an irregular conductor.
- Describe how a lightning rod works.
- Explain how a metal car may protect passengers inside from the dangerous electric fields caused by a downed line touching the car.

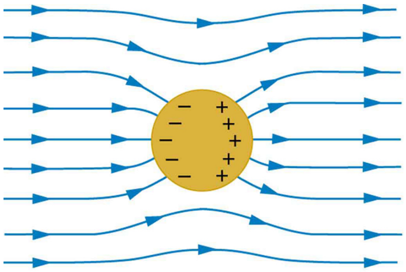
Conductors contain **free charges** that move easily. When excess charge is placed on a conductor or the conductor is put into a static electric field, charges in the conductor quickly respond to reach a steady state called **electrostatic equilibrium**.

[link] shows the effect of an electric field on free charges in a conductor. The free charges move until the field is perpendicular to the conductor's surface. There can be no component of the field parallel to the surface in electrostatic equilibrium, since, if there were, it would produce further movement of charge. A positive free charge is shown, but free charges can be either positive or negative and are, in fact, negative in metals. The motion of a positive charge is equivalent to the motion of a negative charge in the opposite direction.



When an electric field $\mathbf{E}$ is applied to a conductor, free charges inside the conductor move until the field is perpendicular to the surface. (a) The electric field is a vector quantity, with both parallel and perpendicular components. The parallel component ($\mathbf{E}_\parallel$) exerts a force ($\mathbf{F}_\parallel$) on the free charge q, which moves the charge until $\mathbf{F}_\parallel = 0$. (b) The resulting field is perpendicular to the surface. The free charge has been brought to the conductor’s surface, leaving electrostatic forces in equilibrium.

A conductor placed in an **electric field** will be **polarized**. [link] shows the result of placing a neutral conductor in an originally uniform electric field. The field becomes stronger near the conductor but entirely disappears inside it.

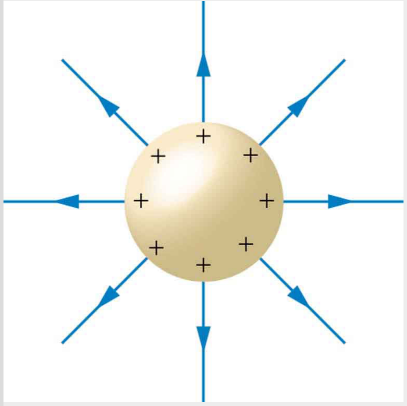


This illustration shows a spherical conductor in static equilibrium with an originally uniform electric field. Free charges move within the conductor, polarizing it, until the electric field lines are perpendicular to the surface. The field lines end on excess negative charge on one section of the surface and begin again on excess positive charge on the opposite side. No electric field exists inside the conductor, since free charges in the conductor would continue moving in response to any field until it was neutralized.

Note:

Misconception Alert: Electric Field inside a Conductor

Excess charges placed on a spherical conductor repel and move until they are evenly distributed, as shown in [link]. Excess charge is forced to the surface until the field inside the conductor is zero. Outside the conductor, the field is exactly the same as if the conductor were replaced by a point charge at its center equal to the excess charge.



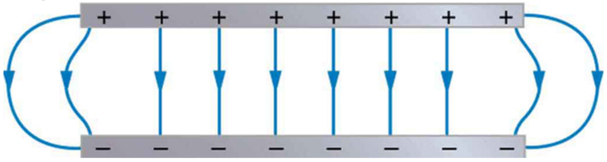
The mutual repulsion of excess positive charges on a spherical conductor distributes them uniformly on its surface. The resulting electric field is perpendicular to the surface and zero inside. Outside the conductor, the field is identical to that of a point charge at the center equal to the excess charge.

Note:
Properties of a Conductor in Electrostatic Equilibrium

1. The electric field is zero inside a conductor.
2. Just outside a conductor, the electric field lines are perpendicular to its surface, ending or beginning on charges on the surface.
3. Any excess charge resides entirely on the surface or surfaces of a conductor.

The properties of a conductor are consistent with the situations already discussed and can be used to analyze any conductor in electrostatic equilibrium. This can lead to some interesting new insights, such as described below.

How can a very uniform electric field be created? Consider a system of two metal plates with opposite charges on them, as shown in [link]. The properties of conductors in electrostatic equilibrium indicate that the electric field between the plates will be uniform in strength and direction. Except near the edges, the excess charges distribute themselves uniformly, producing field lines that are uniformly spaced (hence uniform in strength) and perpendicular to the surfaces (hence uniform in direction, since the plates are flat). The edge effects are less important when the plates are close together.



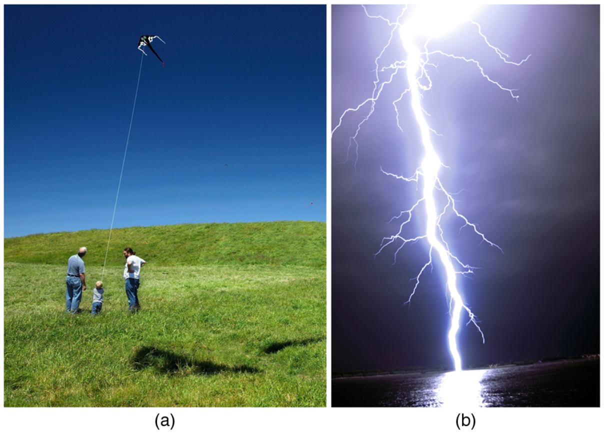
Two metal plates with equal, but opposite, excess charges. The field between them is uniform in strength and direction except near the edges. One use of such a field is to produce uniform acceleration of charges between the plates, such as in the electron gun of a TV tube.

Earth’s Electric Field

A near uniform electric field of approximately 150 N/C, directed downward, surrounds Earth, with the magnitude increasing slightly as we get closer to the surface. What causes the electric field? At around 100 km above the surface of Earth we have a layer of charged particles, called the **ionosphere**. The ionosphere is responsible for a range of phenomena including the electric field surrounding Earth. In fair weather the ionosphere is positive and the Earth largely negative, maintaining the electric field ([link](a)).

In storm conditions clouds form and localized electric fields can be larger and reversed in direction ([link](b)). The exact charge distributions depend on the local conditions, and variations of [link](b) are possible.

If the electric field is sufficiently large, the insulating properties of the surrounding material break down and it becomes conducting. For air this occurs at around 3×10^6 N/C. Air ionizes ions and electrons recombine, and we get discharge in the form of lightning sparks and corona discharge.



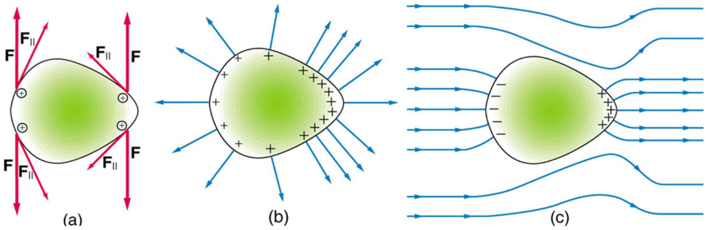
Earth’s electric field. (a) Fair weather field. Earth and the ionosphere (a layer of charged particles) are both conductors. They produce a uniform electric field of about 150 N/C. (credit: D. H. Parks) (b) Storm fields. In the presence of storm clouds, the local electric fields can be larger. At very high fields, the insulating properties of the air break down and lightning can occur. (credit: Jan-Joost Verhoef)

Electric Fields on Uneven Surfaces

So far we have considered excess charges on a smooth, symmetrical conductor surface. What happens if a conductor has sharp corners or is pointed? Excess charges on a nonuniform conductor become concentrated at the sharpest points. Additionally, excess charge may move on or off the conductor at the sharpest points.

To see how and why this happens, consider the charged conductor in [link]. The electrostatic repulsion of like charges is most effective in moving them apart on the flattest surface, and so they become least concentrated there. This is because the forces between identical pairs of charges at either end of the conductor are identical, but the components of the forces parallel to the surfaces are different. The component parallel to the surface is greatest on the flattest surface and, hence, more effective in moving the charge.

The same effect is produced on a conductor by an externally applied electric field, as seen in [link] (c). Since the field lines must be perpendicular to the surface, more of them are concentrated on the most curved parts.



Excess charge on a nonuniform conductor becomes most concentrated at the location of greatest curvature. (a) The forces between identical pairs of charges at either end of the conductor are identical, but the components of the forces parallel to the surface are different. It is $\mathbf{F}_{||}$ that moves the charges apart once they have reached the surface. (b) $\mathbf{F}_{||}$ is smallest at the more pointed end, the charges are left closer together, producing the electric field shown. (c) An uncharged conductor in an originally uniform electric field is polarized, with the most concentrated charge at its most pointed end.

Applications of Conductors

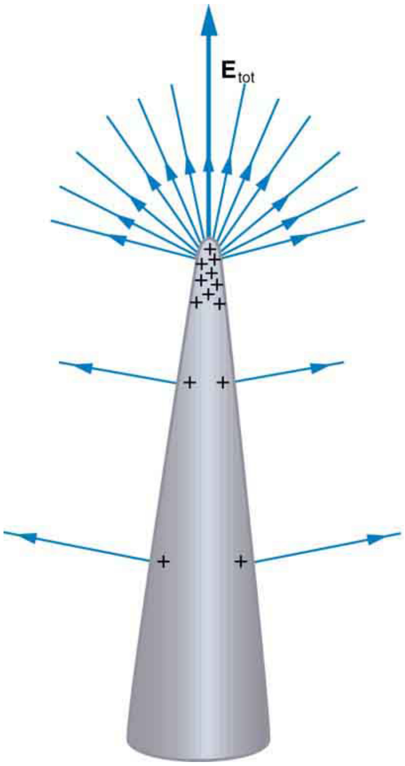
On a very sharply curved surface, such as shown in [link], the charges are so concentrated at the point that the resulting electric field can be great enough to remove them from the surface. This can be useful.

Lightning rods work best when they are most pointed. The large charges created in storm clouds induce an opposite charge on a building that can result in a lightning bolt hitting the building. The induced charge is bled away continually by a lightning rod, preventing the more dramatic lightning strike.

Of course, we sometimes wish to prevent the transfer of charge rather than to facilitate it. In that case, the conductor should be very smooth and have as large a radius of curvature as possible. (See [link].) Smooth surfaces are used on high-voltage transmission lines, for example, to avoid leakage of charge into the air.

Another device that makes use of some of these principles is a **Faraday cage**. This is a metal shield that encloses a volume. All electrical charges will reside on the outside surface of this shield, and there will be no electrical field inside. A Faraday cage is used to prohibit stray electrical fields in the environment from interfering with sensitive measurements, such as the electrical signals inside a nerve cell.

During electrical storms if you are driving a car, it is best to stay inside the car as its metal body acts as a Faraday cage with zero electrical field inside. If in the vicinity of a lightning strike, its effect is felt on the outside of the car and the inside is unaffected, provided you remain totally inside. This is also true if an active (“hot”) electrical wire was broken (in a storm or an accident) and fell on your car.



A very pointed conductor has a large charge concentration at the point. The electric field is very strong at the point and can exert a force large enough to transfer charge on or off the conductor.

Lightning rods are used to prevent the buildup of large excess charges on structures and, thus, are pointed.

(a)

(b)

(a) A lightning rod is pointed to facilitate the transfer of charge. (credit: Romaine, Wikimedia Commons) (b) This Van de Graaff generator has a smooth surface with a large radius of curvature to prevent the transfer of charge and allow a large voltage to be generated. The mutual repulsion of like charges is evident in the person’s hair while touching the metal sphere. (credit: Jon ‘ShakataGaNai’ Davis/Wikimedia Commons).

Section Summary

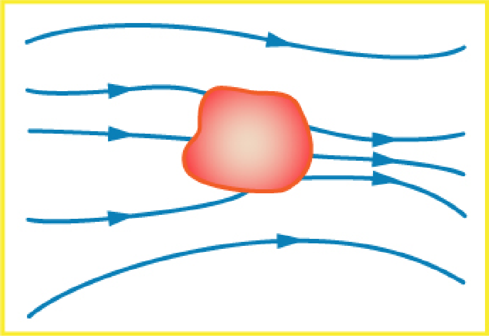
- A conductor allows free charges to move about within it.
- The electrical forces around a conductor will cause free charges to move around inside the conductor until static equilibrium is reached.
- Any excess charge will collect along the surface of a conductor.
- Conductors with sharp corners or points will collect more charge at those points.
- A lightning rod is a conductor with sharply pointed ends that collect excess charge on the building caused by an electrical storm and allow it to dissipate back into the air.
- Electrical storms result when the electrical field of Earth’s surface in certain locations becomes more strongly charged, due to changes in the insulating effect of the air.
- A Faraday cage acts like a shield around an object, preventing electric charge from penetrating inside.

Conceptual Questions

Exercise:

Problem:

Is the object in [link] a conductor or an insulator? Justify your answer.



Exercise:

Problem:

If the electric field lines in the figure above were perpendicular to the object, would it necessarily be a conductor? Explain.

Exercise:

Problem:

The discussion of the electric field between two parallel conducting plates, in this module states that edge effects are less important if the plates are close together. What does close mean? That is, is the actual plate separation crucial, or is the ratio of plate separation to plate area crucial?

Exercise:

Problem:

Would the self-created electric field at the end of a pointed conductor, such as a lightning rod, remove positive or negative charge from the conductor? Would the same sign charge be removed from a neutral pointed conductor by the application of a similar externally created electric field? (The answers to both questions have implications for charge transfer utilizing points.)

Exercise:

Problem:

Why is a golfer with a metal club over her shoulder vulnerable to lightning in an open fairway? Would she be any safer under a tree?

Exercise:

Problem:

Can the belt of a Van de Graaff accelerator be a conductor? Explain.

Exercise:

Problem:

Are you relatively safe from lightning inside an automobile? Give two reasons.

Exercise:

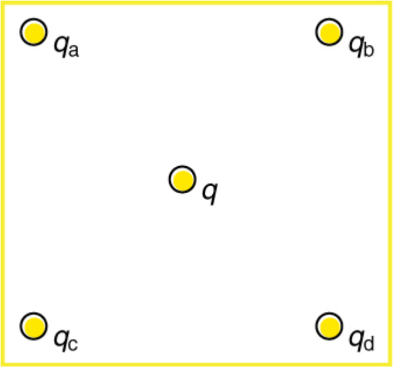
Problem:

Discuss pros and cons of a lightning rod being grounded versus simply being attached to a building.

Exercise:

Problem:

Using the symmetry of the arrangement, show that the net Coulomb force on the charge *q* at the center of the square below ([link]) is zero if the charges on the four corners are exactly equal.



Four point charges q_a, q_b, q_c, and q_d lie on the corners of a square and q is located at its center.

Exercise:

Problem:

(a) Using the symmetry of the arrangement, show that the electric field at the center of the square in [link] is zero if the charges on the four corners are exactly equal. (b) Show that this is also true for any combination of charges in which $q_a = q_d$ and $q_b = q_c$

Exercise:

Problem:

(a) What is the direction of the total Coulomb force on q in [link] if q is negative, $q_a = q_c$ and both are negative, and $q_b = q_c$ and both are positive? (b) What is the direction of the electric field at the center of the square in this situation?

Exercise:

Problem:

Considering [link], suppose that $q_a = q_d$ and $q_b = q_c$. First show that q is in static equilibrium. (You may neglect the gravitational force.) Then discuss whether the equilibrium is stable or unstable, noting that this may depend on the signs of the charges and the direction of displacement of q from the center of the square.

Exercise:

Problem:

If $q_a = 0$ in [link], under what conditions will there be no net Coulomb force on q?

Exercise:

Problem:

In regions of low humidity, one develops a special “grip” when opening car doors, or touching metal door knobs. This involves placing as much of the hand on the device as possible, not just the ends of one’s fingers. Discuss the induced charge and explain why this is done.

Exercise:

Problem:

Tollbooth stations on roadways and bridges usually have a piece of wire stuck in the pavement before them that will touch a car as it approaches. Why is this done?

Exercise:

Problem:

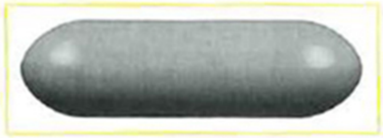
Suppose a woman carries an excess charge. To maintain her charged status can she be standing on ground wearing just any pair of shoes? How would you discharge her? What are the consequences if she simply walks away?

Problems & Exercises

Exercise:

Problem:

Sketch the electric field lines in the vicinity of the conductor in [link] given the field was originally uniform and parallel to the object's long axis. Is the resulting field small near the long side of the object?



Exercise:

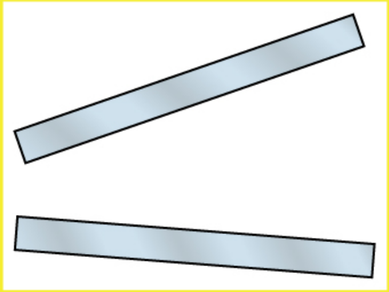
Problem:

Sketch the electric field lines in the vicinity of the conductor in [link] given the field was originally uniform and parallel to the object's long axis. Is the resulting field small near the long side of the object?

Exercise:

Problem:

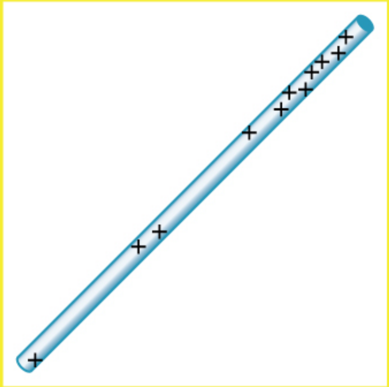
Sketch the electric field between the two conducting plates shown in [link], given the top plate is positive and an equal amount of negative charge is on the bottom plate. Be certain to indicate the distribution of charge on the plates.



Exercise:

Problem:

Sketch the electric field lines in the vicinity of the charged insulator in [link] noting its nonuniform charge distribution.



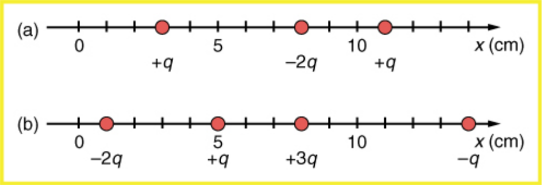
A charged
insulating rod such
as might be used in

a classroom demonstration.

Exercise:

Problem:

What is the force on the charge located at $x = 8.00 \text{ cm}$ in [link](a) given that $q = 1.00 \text{ μC}$?



(a) Point charges located at 3.00, 8.00, and 11.0 cm along the *x*-axis. (b) Point charges located at 1.00, 5.00, 8.00, and 14.0 cm along the *x*-axis.

Exercise:

Problem:

(a) Find the total electric field at $x = 1.00 \text{ cm}$ in [link](b) given that $q = 5.00 \text{ nC}$. (b) Find the total electric field at $x = 11.00 \text{ cm}$ in [link](b). (c) If the charges are allowed to move and eventually be brought to rest by friction, what will the final charge configuration be? (That is, will there be a single charge, double charge, etc., and what will its value(s) be?)

Solution:

(a) $E_{x=1.00 \text{ cm}} = -\infty$

(b) $2.12 \times 10^5 \text{ N/C}$

(c) one charge of $+q$

Exercise:

Problem:

(a) Find the electric field at $x = 5.00 \text{ cm}$ in [link](a), given that $q = 1.00 \text{ µC}$. (b) At what position between 3.00 and 8.00 cm is the total electric field the same as that for $–2q$ alone? (c) Can the electric field be zero anywhere between 0.00 and 8.00 cm? (d) At very large positive or negative values of *x*, the electric field approaches zero in both (a) and (b). In which does it most rapidly approach zero and why? (e) At what position to the right of 11.0 cm is the total electric field zero, other than at infinity? (Hint: A graphing calculator can yield considerable insight in this problem.)

Exercise:

Problem:

(a) Find the total Coulomb force on a charge of 2.00 nC located at $x = 4.00 \text{ cm}$ in [link] (b), given that $q = 1.00 \text{ µC}$. (b) Find the *x*-position at which the electric field is zero in [link] (b).

Solution:

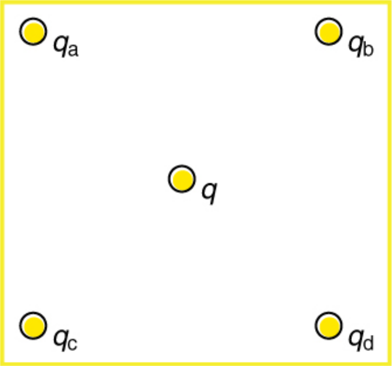
(a) 0.252 N to the left

(b) $x = 6.07 \text{ cm}$

Exercise:

Problem:

Using the symmetry of the arrangement, determine the direction of the force on q in the figure below, given that $q_a = q_b$=$+7.50 \text{ µC}$ and $q_c = q_d = -7.50 \text{ µC}$. (b) Calculate the magnitude of the force on the charge q, given that the square is 10.0 cm on a side and $q = 2.00 \text{ µC}$.



Exercise:

Problem:

(a) Using the symmetry of the arrangement, determine the direction of the electric field at the center of the square in [link], given that $q_a = q_b = -1.00\ \mu\text{C}$ and $q_c = q_d$=$+1.00\ \mu\text{C}$. (b) Calculate the magnitude of the electric field at the location of q, given that the square is 5.00 cm on a side.

Solution:

(a)The electric field at the center of the square will be straight up, since q_a and q_b are positive and q_c and q_d are negative and all have the same magnitude.

(b) $2.04 \times 10^7\ \text{N/C (upward)}$

Exercise:

Problem:

Find the electric field at the location of q_a in [link] given that $q_b = q_c = q_d$=$+2.00\ \text{nC}$, $q = -1.00\ \text{nC}$, and the square is 20.0 cm on a side.

Exercise:

Problem:

Find the total Coulomb force on the charge q in [link], given that $q = 1.00\ \mu\text{C}$, $q_a = 2.00\ \mu\text{C}$, $q_b = -3.00\ \mu\text{C}$, $q_c = -4.00\ \mu\text{C}$, and $q_d = +1.00\ \mu\text{C}$. The square is 50.0 cm on a side.

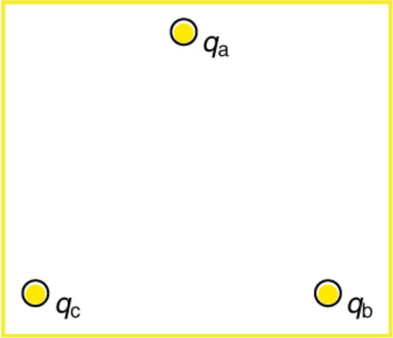
Solution:

$0.102\ \text{N}$, in the $-y$ direction

Exercise:

Problem:

(a) Find the electric field at the location of q_a in [link], given that $q_\text{b} = +10.00\ \mu\text{C}$ and $q_\text{c} = -5.00\ \mu\text{C}$. (b) What is the force on q_a, given that $q_\text{a} = +1.50\ \text{nC}$?



Point charges located at the corners of an equilateral triangle 25.0 cm on a side.

Exercise:

Problem:

(a) Find the electric field at the center of the triangular configuration of charges in [link], given that $q_a = +2.50\ \text{nC}$, $q_b = -8.00\ \text{nC}$, and $q_c = +1.50\ \text{nC}$. (b) Is there any combination of charges, other than $q_a = q_b = q_c$, that will produce a zero strength electric field at the center of the triangular configuration?

Solution:

(a) $\vec{E} = 4.36 \times 10^3\ \text{N/C}, 35.0^\text{o}$, below the horizontal.

(b) No

Glossary

conductor	an object with properties that allow charges to move about freely within it
free charge	an electrical charge (either positive or negative) which can move about separately from its base molecule
electrostatic equilibrium	an electrostatically balanced state in which all free electrical charges have stopped moving about
polarized	a state in which the positive and negative charges within an object have collected in separate locations
ionosphere	a layer of charged particles located around 100 km above the surface of Earth, which is responsible for a range of phenomena including the electric field surrounding Earth

Faraday cage

a metal shield which prevents electric charge from penetrating its surface

Applications of Electrostatics

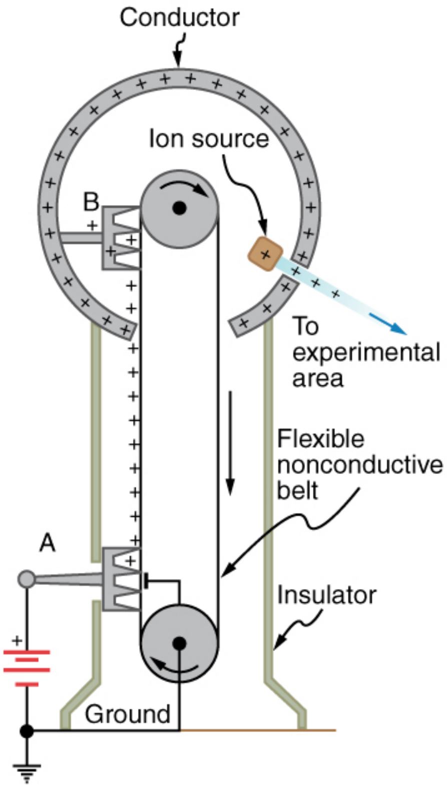
- Name several real-world applications of the study of electrostatics.

The study of **electrostatics** has proven useful in many areas. This module covers just a few of the many applications of electrostatics.

The Van de Graaff Generator

Van de Graaff generators (or Van de Graaffs) are not only spectacular devices used to demonstrate high voltage due to static electricity—they are also used for serious research. The first was built by Robert Van de Graaff in 1931 (based on original suggestions by Lord Kelvin) for use in nuclear physics research. [link] shows a schematic of a large research version. Van de Graaffs utilize both smooth and pointed surfaces, and conductors and insulators to generate large static charges and, hence, large voltages.

A very large excess charge can be deposited on the sphere, because it moves quickly to the outer surface. Practical limits arise because the large electric fields polarize and eventually ionize surrounding materials, creating free charges that neutralize excess charge or allow it to escape. Nevertheless, voltages of 15 million volts are well within practical limits.


Conductor
Ion source
B
To experimental area
Flexible nonconductive belt
A
Insulator
Ground

Schematic of Van de Graaff generator. A battery (A) supplies excess positive charge to a pointed conductor, the points of which spray the charge onto a moving insulating belt near the bottom. The pointed conductor (B) on top in the large sphere picks up the charge. (The induced electric field at the points is so large that it removes the charge from the belt.) This can be done because the charge does not remain inside the conducting sphere but moves to its outside surface. An ion source inside the sphere produces positive ions, which are accelerated away from the positive sphere to high velocities.

Note:

Take-Home Experiment: Electrostatics and Humidity

Rub a comb through your hair and use it to lift pieces of paper. It may help to tear the pieces of paper rather than cut them neatly. Repeat the exercise in your bathroom after you have had a long shower and the air in the bathroom is moist. Is it easier to get electrostatic effects in dry or moist air? Why would torn paper be more attractive to the comb than cut paper? Explain your observations.

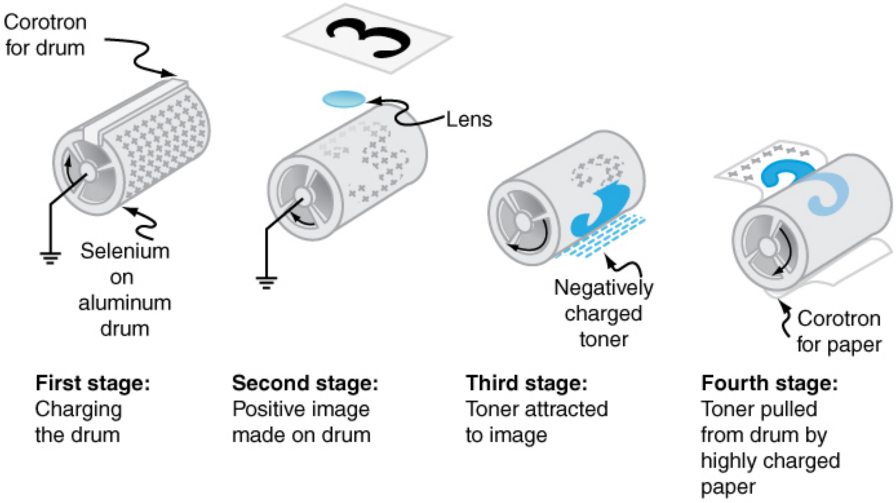
Xerography

Most copy machines use an electrostatic process called **xerography**—a word coined from the Greek words *xeros* for dry and *graphos* for writing. The heart of the process is shown in simplified form in [link].

A selenium-coated aluminum drum is sprayed with positive charge from points on a device called a corotron. Selenium is a substance with an interesting property—it is a **photoconductor**. That is, selenium is an insulator when in the dark and a conductor when exposed to light.

In the first stage of the xerography process, the conducting aluminum drum is **grounded** so that a negative charge is induced under the thin layer of uniformly positively charged selenium. In the second stage, the surface of the drum is exposed to the image of whatever is to be copied. Where the image is light, the selenium becomes conducting, and the positive charge is neutralized. In dark areas, the positive charge remains, and so the image has been transferred to the drum.

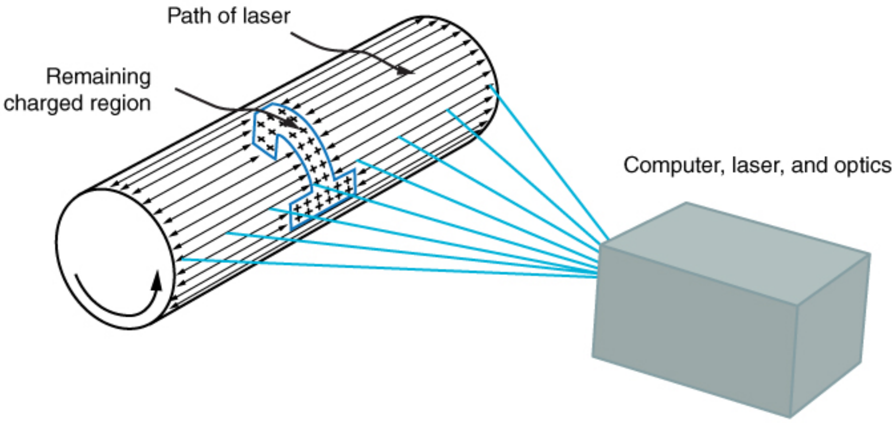
The third stage takes a dry black powder, called toner, and sprays it with a negative charge so that it will be attracted to the positive regions of the drum. Next, a blank piece of paper is given a greater positive charge than on the drum so that it will pull the toner from the drum. Finally, the paper and electrostatically held toner are passed through heated pressure rollers, which melt and permanently adhere the toner within the fibers of the paper.


Corotron for drum
Selenium on aluminum drum
Lens
Negatively charged toner
Corotron for paper
First stage: Charging the drum
Second stage: Positive image made on drum
Third stage: Toner attracted to image
Fourth stage: Toner pulled from drum by highly charged paper

Xerography is a dry copying process based on electrostatics. The major steps in the process are the charging of the photoconducting drum, transfer of an image creating a positive charge duplicate, attraction of toner to the charged parts of the drum, and transfer of toner to the paper. Not shown are heat treatment of the paper and cleansing of the drum for the next copy.

Laser Printers

Laser printers use the xerographic process to make high-quality images on paper, employing a laser to produce an image on the photoconducting drum as shown in [link]. In its most common application, the laser printer receives output from a computer, and it can achieve high-quality output because of the precision with which laser light can be controlled. Many laser printers do significant information processing, such as making sophisticated letters or fonts, and may contain a computer more powerful than the one giving them the raw data to be printed.

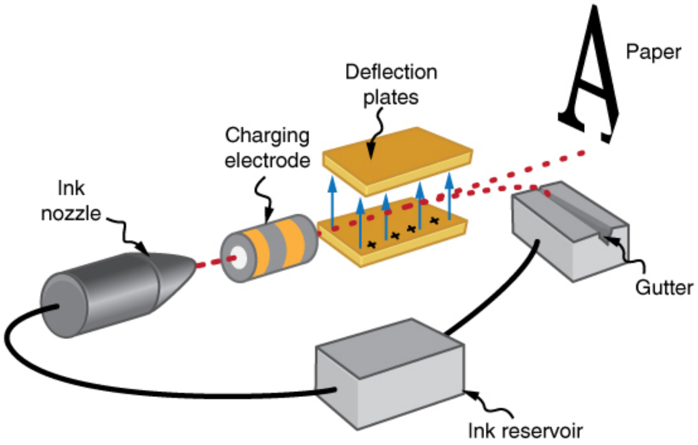


In a laser printer, a laser beam is scanned across a photoconducting drum, leaving a positive charge image. The other steps for charging the drum and transferring the image to paper are the same as in xerography. Laser light can be very precisely controlled, enabling laser printers to produce high-quality images.

Ink Jet Printers and Electrostatic Painting

The **ink jet printer**, commonly used to print computer-generated text and graphics, also employs electrostatics. A nozzle makes a fine spray of tiny ink droplets, which are then given an electrostatic charge. (See [link].)

Once charged, the droplets can be directed, using pairs of charged plates, with great precision to form letters and images on paper. Ink jet printers can produce color images by using a black jet and three other jets with primary colors, usually cyan, magenta, and yellow, much as a color television produces color. (This is more difficult with xerography, requiring multiple drums and toners.)



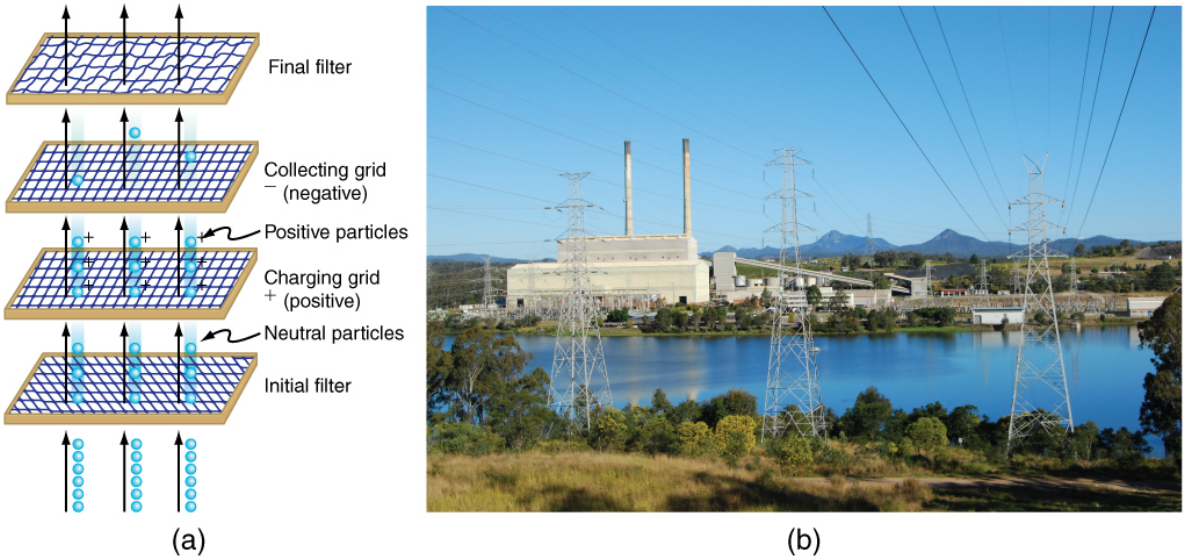
The nozzle of an ink-jet printer produces small ink droplets, which are sprayed with electrostatic charge. Various computer-driven devices are then used to direct the droplets to the correct positions on a page.

Electrostatic painting employs electrostatic charge to spray paint onto odd-shaped surfaces. Mutual repulsion of like charges causes the paint to fly away from its source. Surface tension forms drops, which are then attracted by unlike charges to the surface to be painted. Electrostatic painting can reach those hard-to-get at places, applying an even coat in a controlled manner. If the object is a conductor, the electric field is perpendicular to the surface, tending to bring the drops in perpendicularly. Corners and points on conductors will receive extra paint. Felt can similarly be applied.

Smoke Precipitators and Electrostatic Air Cleaning

Another important application of electrostatics is found in air cleaners, both large and small. The electrostatic part of the process places excess (usually positive) charge on smoke, dust, pollen, and other particles in the air and then passes the air through an oppositely charged grid that attracts and retains the charged particles. (See [link].)

Large **electrostatic precipitators** are used industrially to remove over 99% of the particles from stack gas emissions associated with the burning of coal and oil. Home precipitators, often in conjunction with the home heating and air conditioning system, are very effective in removing polluting particles, irritants, and allergens.



(a) Schematic of an electrostatic precipitator. Air is passed through grids of opposite charge. The first grid charges airborne particles, while the second attracts and collects them. (b) The dramatic effect of electrostatic precipitators is seen by the absence of smoke from this power plant. (credit: Cmdalgleish, Wikimedia Commons)

Note:
Problem-Solving Strategies for Electrostatics

1. Examine the situation to determine if static electricity is involved. This may concern separated stationary charges, the forces among them, and the electric fields they create.
2. Identify the system of interest. This includes noting the number, locations, and types of charges involved.
3. Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful. Determine whether the Coulomb force is to be considered directly—if so, it may be useful to draw a free-body diagram, using electric field lines.
4. Make a list of what is given or can be inferred from the problem as stated (identify the knowns). It is important to distinguish the Coulomb force F from the electric field E, for example.
5. Solve the appropriate equation for the quantity to be determined (the unknown) or draw the field lines as requested.
6. Examine the answer to see if it is reasonable: Does it make sense? Are units correct and the numbers involved reasonable?

Integrated Concepts

The Integrated Concepts exercises for this module involve concepts such as electric charges, electric fields, and several other topics. Physics is most interesting when applied to general situations involving more than a narrow set of physical principles. The electric field exerts force on charges, for example, and hence the relevance of Dynamics: Force and Newton’s Laws of Motion. The following topics are involved in some or all of the problems labeled “Integrated Concepts”:

- Kinematics
- Two-Dimensional Kinematics
- Dynamics: Force and Newton’s Laws of Motion
- Uniform Circular Motion and Gravitation
- Statics and Torque
- Fluid Statics

The following worked example illustrates how this strategy is applied to an Integrated Concept problem:

Example:

Acceleration of a Charged Drop of Gasoline

If steps are not taken to ground a gasoline pump, static electricity can be placed on gasoline when filling your car’s tank. Suppose a tiny drop of gasoline has a mass of 4.00×10^{-15} kg and is given a positive charge of $3.20 \times 10^{-19}\ \text{C}$. (a) Find the weight of the drop. (b) Calculate the electric force on the drop if there is an upward electric field of strength $3.00 \times 10^{5}\ \text{N/C}$ due to other static electricity in the vicinity. (c) Calculate the drop’s acceleration.

Strategy

To solve an integrated concept problem, we must first identify the physical principles involved and identify the chapters in which they are found. Part (a) of this example asks for weight. This is a topic of dynamics and is defined in Dynamics: Force and Newton’s Laws of Motion. Part (b) deals with electric force on a charge, a topic of Electric Charge and Electric Field. Part (c) asks for acceleration, knowing forces and mass. These are part of Newton’s laws, also found in Dynamics: Force and Newton’s Laws of Motion.

The following solutions to each part of the example illustrate how the specific problem-solving strategies are applied. These involve identifying knowns and unknowns, checking to see if the answer is reasonable, and so on.

Solution for (a)

Weight is mass times the acceleration due to gravity, as first expressed in

Equation:

$$w = \text{mg}.$$

Entering the given mass and the average acceleration due to gravity yields

Equation:

$$w = (4.00 \times 10^{-15}\ \text{kg})(9.80\ \text{m/s}^2) = 3.92 \times 10^{-14}\ \text{N}.$$

Discussion for (a)

This is a small weight, consistent with the small mass of the drop.

Solution for (b)

The force an electric field exerts on a charge is given by rearranging the following equation:

Equation:

$$F = \text{qE}.$$

Here we are given the charge (3.20×10^{-19} C is twice the fundamental unit of charge) and the electric field strength, and so the electric force is found to be

Equation:

$$F = (3.20 \times 10^{-19}\ \text{C})(3.00 \times 10^{5}\ \text{N/C}) = 9.60 \times 10^{-14}\ \text{N}.$$

Discussion for (b)

While this is a small force, it is greater than the weight of the drop.

Solution for (c)

The acceleration can be found using Newton’s second law, provided we can identify all of the external forces acting on the drop. We assume only the drop’s weight and the electric force are significant. Since the drop has a positive charge and the electric field is given to be upward, the electric force is upward. We thus have a one-dimensional (vertical direction) problem, and we can state Newton’s second law as

Equation:

$$a = \frac{F_{\text{net}}}{m}.$$

where $F_{\text{net}} = F - w$. Entering this and the known values into the expression for Newton’s second law yields

Equation:

$$\begin{aligned} a &= \frac{F-w}{m} \\ &= \frac{9.60\times10^{-14}\text{ N}-3.92\times10^{-14}\text{ N}}{4.00\times10^{-15}\text{ kg}} \\ &= 14.2\text{ m/s}^2. \end{aligned}$$

Discussion for (c)

This is an upward acceleration great enough to carry the drop to places where you might not wish to have gasoline.

This worked example illustrates how to apply problem-solving strategies to situations that include topics in different chapters. The first step is to identify the physical principles involved in the problem. The second step is to solve for the unknown using familiar problem-solving strategies. These are found throughout the text, and many worked examples show how to use them for single topics. In this integrated concepts example, you can see how to apply them across several topics. You will find these techniques useful in applications of physics outside a physics course, such as in your profession, in other science disciplines, and in everyday life. The following problems will build your skills in the broad application of physical principles.

Note:

Unreasonable Results

The Unreasonable Results exercises for this module have results that are unreasonable because some premise is unreasonable or because certain of the premises are inconsistent with one another. Physical principles applied correctly then produce unreasonable results. The purpose of these problems is to give practice in assessing whether nature is being accurately described, and if it is not to trace the source of difficulty.

Note:

Problem-Solving Strategy

To determine if an answer is reasonable, and to determine the cause if it is not, do the following.

1. Solve the problem using strategies as outlined above. Use the format followed in the worked examples in the text to solve the problem as usual.
2. Check to see if the answer is reasonable. Is it too large or too small, or does it have the wrong sign, improper units, and so on?
3. If the answer is unreasonable, look for what specifically could cause the identified difficulty. Usually, the manner in which the answer is unreasonable is an indication of the difficulty. For example, an extremely large Coulomb force could be due to the assumption of an excessively large separated charge.

Section Summary

- Electrostatics is the study of electric fields in static equilibrium.
- In addition to research using equipment such as a Van de Graaff generator, many practical applications of electrostatics exist, including photocopiers, laser printers, ink-jet printers and electrostatic air filters.

Problems & Exercises

Exercise:

Problem:

(a) What is the electric field 5.00 m from the center of the terminal of a Van de Graaff with a 3.00 mC charge, noting that the field is equivalent to that of a point charge at the center of the terminal? (b) At this distance, what force does the field exert on a $2.00\ \mu\text{C}$ charge on the Van de Graaff's belt?

Exercise:

Problem:

(a) What is the direction and magnitude of an electric field that supports the weight of a free electron near the surface of Earth? (b) Discuss what the small value for this field implies regarding the relative strength of the gravitational and electrostatic forces.

Solution:

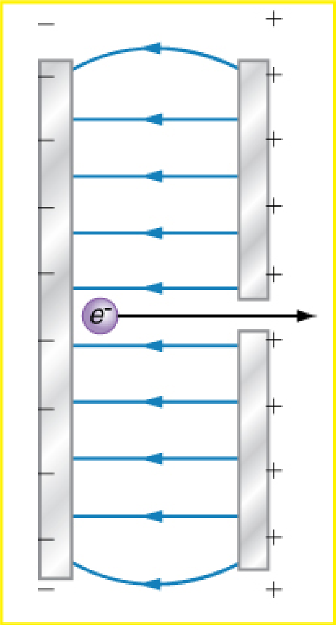
(a) $5.58 \times 10^{-11}\ \text{N/C}$

(b)the coulomb force is extraordinarily stronger than gravity

Exercise:

Problem:

A simple and common technique for accelerating electrons is shown in [link], where there is a uniform electric field between two plates. Electrons are released, usually from a hot filament, near the negative plate, and there is a small hole in the positive plate that allows the electrons to continue moving. (a) Calculate the acceleration of the electron if the field strength is $2.50 \times 10^{4}\ \text{N/C}$. (b) Explain why the electron will not be pulled back to the positive plate once it moves through the hole.



Parallel conducting plates with opposite charges on them create a relatively uniform electric field used to accelerate electrons to the right. Those that go through the hole can be used to make a TV or computer screen glow or to produce X-rays.

Exercise:

Problem:

Earth has a net charge that produces an electric field of approximately 150 N/C downward at its surface. (a) What is the magnitude and sign of the excess charge, noting the electric field of a conducting sphere is equivalent to a point charge at its center? (b) What acceleration will the field produce on a free electron near Earth's surface? (c) What mass object with a single extra electron will have its weight supported by this field?

Solution:

(a) $-6.76 \times 10^5 \text{ C}$

(b) $2.63 \times 10^{13} \text{ m/s}^2 \text{ (upward)}$

(c) $2.45 \times 10^{-18} \text{ kg}$

Exercise:

Problem:

Point charges of $25.0\ \mu\text{C}$ and $45.0\ \mu\text{C}$ are placed 0.500 m apart. (a) At what point along the line between them is the electric field zero? (b) What is the electric field halfway between them?

Exercise:

Problem:

What can you say about two charges q_1 and q_2, if the electric field one-fourth of the way from q_1 to q_2 is zero?

Solution:

The charge q_2 is 9 times greater than q_1.

Exercise:

Problem: **Integrated Concepts**

Calculate the angular velocity ω of an electron orbiting a proton in the hydrogen atom, given the radius of the orbit is $0.530 \times 10^{-10}\ \text{m}$. You may assume that the proton is stationary and the centripetal force is supplied by Coulomb attraction.

Exercise:

Problem: Integrated Concepts

An electron has an initial velocity of $5.00 \times 10^{6}\ \text{m/s}$ in a uniform $2.00 \times 10^{5}\ \text{N/C}$ strength electric field. The field accelerates the electron in the direction opposite to its initial velocity. (a) What is the direction of the electric field? (b) How far does the electron travel before coming to rest? (c) How long does it take the electron to come to rest? (d) What is the electron's velocity when it returns to its starting point?

Exercise:

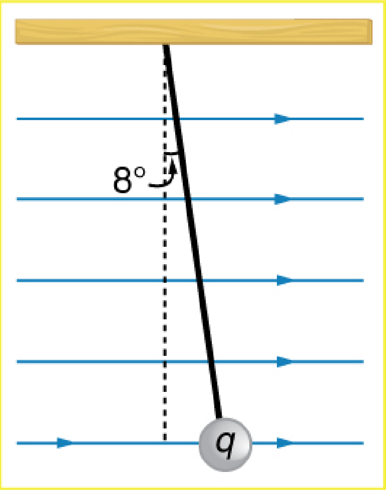
Problem: Integrated Concepts

The practical limit to an electric field in air is about $3.00 \times 10^{6}\ \text{N/C}$. Above this strength, sparking takes place because air begins to ionize and charges flow, reducing the field. (a) Calculate the distance a free proton must travel in this field to reach 3.00% of the speed of light, starting from rest. (b) Is this practical in air, or must it occur in a vacuum?

Exercise:

Problem: Integrated Concepts

A 5.00 g charged insulating ball hangs on a 30.0 cm long string in a uniform horizontal electric field as shown in [link]. Given the charge on the ball is $1.00\ \mu\text{C}$, find the strength of the field.

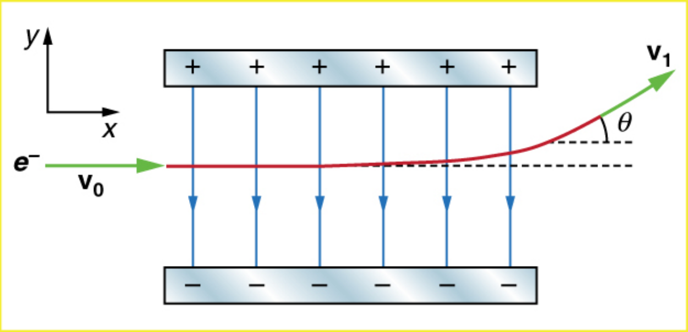


A horizontal electric field causes the charged ball to hang at an angle of 8.00º.

Exercise:

Problem: Integrated Concepts

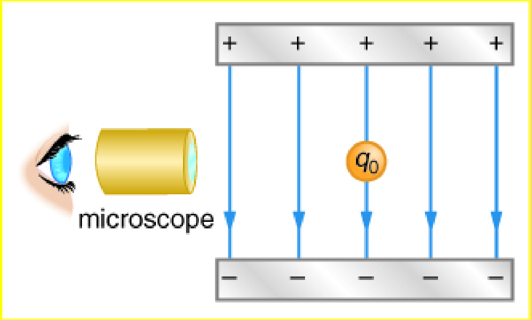
[link] shows an electron passing between two charged metal plates that create an 100 N/C vertical electric field perpendicular to the electron’s original horizontal velocity. (These can be used to change the electron’s direction, such as in an oscilloscope.) The initial speed of the electron is $3.00 \times 10^6 \text{ m/s}$, and the horizontal distance it travels in the uniform field is 4.00 cm. (a) What is its vertical deflection? (b) What is the vertical component of its final velocity? (c) At what angle does it exit? Neglect any edge effects.



Exercise:

Problem: Integrated Concepts

The classic Millikan oil drop experiment was the first to obtain an accurate measurement of the charge on an electron. In it, oil drops were suspended against the gravitational force by a vertical electric field. (See [link].) Given the oil drop to be $1.00\ \mu\text{m}$ in radius and have a density of $920\ \text{kg/m}^3$: (a) Find the weight of the drop. (b) If the drop has a single excess electron, find the electric field strength needed to balance its weight.



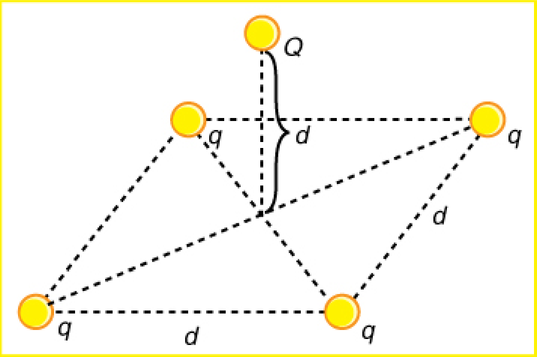
In the Millikan oil drop experiment, small drops can be suspended in an electric field by the force exerted on a single excess electron. Classically, this experiment was used to determine the electron charge q_e by

measuring the electric field
and mass of the drop.

Exercise:

Problem: **Integrated Concepts**

(a) In [link], four equal charges q lie on the corners of a square. A fifth charge Q is on a mass m directly above the center of the square, at a height equal to the length d of one side of the square. Determine the magnitude of q in terms of Q, m, and d, if the Coulomb force is to equal the weight of m. (b) Is this equilibrium stable or unstable? Discuss.



Four equal charges on the corners of a horizontal square support the weight of a fifth charge located directly above the center of the square.

Exercise:

Problem: **Unreasonable Results**

(a) Calculate the electric field strength near a 10.0 cm diameter conducting sphere that has 1.00 C of excess charge on it. (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem: Unreasonable Results

(a) Two 0.500 g raindrops in a thunderhead are 1.00 cm apart when they each acquire 1.00 mC charges. Find their acceleration. (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Exercise:

Problem: Unreasonable Results

A wrecking yard inventor wants to pick up cars by charging a 0.400 m diameter ball and inducing an equal and opposite charge on the car. If a car has a 1000 kg mass and the ball is to be able to lift it from a distance of 1.00 m: (a) What minimum charge must be used? (b) What is the electric field near the surface of the ball? (c) Why are these results unreasonable? (d) Which premise or assumption is responsible?

Exercise:

Problem: Construct Your Own Problem

Consider two insulating balls with evenly distributed equal and opposite charges on their surfaces, held with a certain distance between the centers of the balls. Construct a problem in which you calculate the electric field (magnitude and direction) due to the balls at various points along a line running through the centers of the balls and extending to infinity on either side. Choose interesting points and comment on the meaning of the field at those points. For example, at what points might the field be just that due to one ball and where does the field become negligibly small? Among the things to be considered are the magnitudes of the charges and the distance between the centers of the balls. Your instructor may wish for you to consider the electric field off axis or for a more complex array of charges, such as those in a water molecule.

Exercise:

Problem: Construct Your Own Problem

Consider identical spherical conducting space ships in deep space where gravitational fields from other bodies are negligible compared to the gravitational attraction between the ships. Construct a problem in which you place identical excess charges on the space ships to exactly counter their gravitational attraction. Calculate the amount of excess charge needed. Examine whether that charge depends on the distance between the centers of the ships, the masses of the ships, or any other factors. Discuss whether this would be an easy, difficult, or even impossible thing to do in practice.

Glossary

Van de Graaff generator
a machine that produces a large amount of excess charge, used for experiments with high voltage

electrostatics
the study of electric forces that are static or slow-moving

photoconductor
a substance that is an insulator until it is exposed to light, when it becomes a conductor

xerography
a dry copying process based on electrostatics

grounded
connected to the ground with a conductor, so that charge flows freely to and from the Earth to the grounded object

laser printer
uses a laser to create a photoconductive image on a drum, which attracts dry ink particles that are then rolled onto a sheet of paper to print a high-quality copy of the image

ink-jet printer
small ink droplets sprayed with an electric charge are controlled by electrostatic plates to create images on paper

electrostatic precipitators
filters that apply charges to particles in the air, then attract those charges to a filter, removing them from the airstream

Introduction to Electric Potential and Electric Energy class="introduction"

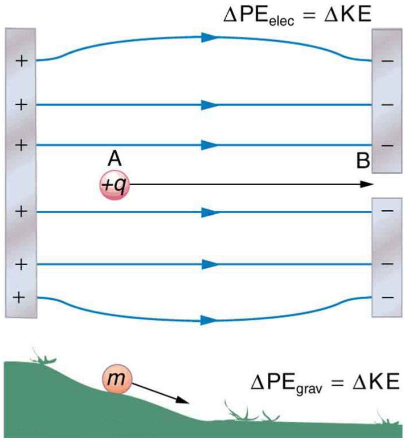
Automated external defibrillator unit (AED) (credit: U.S. Defense Department photo/Tech. Sgt. Suzanne M. Day)

In Electric Charge and Electric Field, we just scratched the surface (or at least rubbed it) of electrical phenomena. Two of the most familiar aspects of electricity are its energy and *voltage*. We know, for example, that great amounts of electrical energy can be stored in batteries, are transmitted cross-country through power lines, and may jump from clouds to explode the sap of trees. In a similar manner, at molecular levels, *ions* cross cell membranes and transfer information. We also know about voltages associated with electricity. Batteries are typically a few volts, the outlets in your home produce 120 volts, and power lines can be as high as hundreds of thousands of volts. But energy and voltage are not the same thing. A motorcycle battery, for example, is small and would not be very successful in replacing the much larger car battery, yet each has the same voltage. In this chapter, we shall examine the relationship between voltage and electrical energy and begin to explore some of the many applications of electricity.

Electric Potential Energy: Potential Difference

- Define electric potential and electric potential energy.
- Describe the relationship between potential difference and electrical potential energy.
- Explain electron volt and its usage in submicroscopic process.
- Determine electric potential energy given potential difference and amount of charge.

When a free positive charge *q* is accelerated by an electric field, such as shown in [link], it is given kinetic energy. The process is analogous to an object being accelerated by a gravitational field. It is as if the charge is going down an electrical hill where its electric potential energy is converted to kinetic energy. Let us explore the work done on a charge *q* by the electric field in this process, so that we may develop a definition of electric potential energy.


$\Delta PE_{elec} = \Delta KE$
+
−
A
B
$+q$
m
$\Delta PE_{grav} = \Delta KE$

A charge accelerated by an electric field is analogous to a mass going down a hill. In both cases potential energy is converted to another form. Work is done by a force, but since this force

is conservative, we
can write
$W = -\Delta \text{PE}.$

The electrostatic or Coulomb force is conservative, which means that the work done on q is independent of the path taken. This is exactly analogous to the gravitational force in the absence of dissipative forces such as friction. When a force is conservative, it is possible to define a potential energy associated with the force, and it is usually easier to deal with the potential energy (because it depends only on position) than to calculate the work directly.

We use the letters PE to denote electric potential energy, which has units of joules (J). The change in potential energy, ΔPE, is crucial, since the work done by a conservative force is the negative of the change in potential energy; that is, $W = -\Delta \text{PE}$. For example, work W done to accelerate a positive charge from rest is positive and results from a loss in PE, or a negative ΔPE. There must be a minus sign in front of ΔPE to make W positive. PE can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Note:
Potential Energy
$W = -\Delta \text{PE}$. For example, work W done to accelerate a positive charge from rest is positive and results from a loss in PE, or a negative ΔPE. There must be a minus sign in front of ΔPE to make W positive. PE can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Gravitational potential energy and electric potential energy are quite analogous. Potential energy accounts for work done by a conservative force and gives added insight regarding energy and energy transformation without the necessity of dealing with the force directly. It is much more common, for example, to use the concept of voltage (related to electric potential energy) than to deal with the Coulomb force directly.

Calculating the work directly is generally difficult, since $W = \text{Fd cos }\theta$ and the direction and magnitude of F can be complex for multiple charges, for odd-shaped objects, and along arbitrary paths. But we do know that, since $F = \text{qE}$, the work, and hence ΔPE, is proportional to the test charge q. To have a physical quantity that is independent of test charge, we define **electric potential** V (or simply potential, since electric is understood) to be the potential energy per unit charge:

Equation:

$$V = \frac{\text{PE}}{q}.$$

Note:

Electric Potential

This is the electric potential energy per unit charge.

Equation:

$$V = \frac{\text{PE}}{q}$$

Since PE is proportional to q , the dependence on q cancels. Thus V does not depend on q. The change in potential energy ΔPE is crucial, and so we are concerned with the difference in potential or potential difference ΔV between two points, where

Equation:

$$\Delta V = V_{\text{B}} - V_{\text{A}} = \frac{\Delta\text{PE}}{q}.$$

The **potential difference** between points A and B, $V_{\text{B}} - V_{\text{A}}$, is thus defined to be the change in potential energy of a charge q moved from A to B, divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.

Equation:

$$1\text{ V} = 1\ \frac{\text{J}}{\text{C}}$$

Note:

Potential Difference

The potential difference between points A and B, $V_{\text{B}} - V_{\text{A}}$, is defined to be the change in potential energy of a charge q moved from A to B, divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.

Equation:

$$1\text{ V} = 1\ \frac{\text{J}}{\text{C}}$$

The familiar term **voltage** is the common name for potential difference. Keep in mind that whenever a voltage is quoted, it is understood to be the potential difference between two points. For example, every battery has two terminals, and its voltage is the potential difference between them. More fundamentally, the point you choose to be zero volts is arbitrary. This is analogous to the fact that gravitational potential energy has an arbitrary zero, such as sea level or perhaps a lecture hall floor.

In summary, the relationship between potential difference (or voltage) and electrical potential energy is given by

Equation:

$$\Delta V = \frac{\Delta \text{PE}}{q} \text{ and } \Delta \text{PE} = q\Delta V.$$

Note:
Potential Difference and Electrical Potential Energy
The relationship between potential difference (or voltage) and electrical potential energy is given by
Equation:

$$\Delta V = \frac{\Delta \text{PE}}{q} \text{ and } \Delta \text{PE} = q\Delta V.$$

The second equation is equivalent to the first.

Voltage is not the same as energy. Voltage is the energy per unit charge. Thus a motorcycle battery and a car battery can both have the same voltage (more precisely, the same potential difference between battery terminals), yet one stores much more energy than the other since $\Delta \text{PE} = q\Delta V$. The car battery can move more charge than the motorcycle battery, although both are 12 V batteries.

Example:
Calculating Energy
Suppose you have a 12.0 V motorcycle battery that can move 5000 C of charge, and a 12.0 V car battery that can move 60,000 C of charge. How much energy does each deliver? (Assume that the numerical value of each charge is accurate to three significant figures.)
Strategy
To say we have a 12.0 V battery means that its terminals have a 12.0 V potential difference. When such a battery moves charge, it puts the charge

through a potential difference of 12.0 V, and the charge is given a change in potential energy equal to $\Delta\text{PE} = q\Delta V$.

So to find the energy output, we multiply the charge moved by the potential difference.

Solution

For the motorcycle battery, $q = 5000\text{ C}$ and $\Delta V = 12.0\text{ V}$. The total energy delivered by the motorcycle battery is

Equation:

$$\begin{array}{rcl}\Delta\text{PE}_{\text{cycle}} & = & (5000\text{ C})(12.0\text{ V})\\ & = & (5000\text{ C})(12.0\text{ J/C})\\ & = & 6.00 \times 10^{4}\text{ J.}\end{array}$$

Similarly, for the car battery, $q = 60{,}000\text{ C}$ and

Equation:

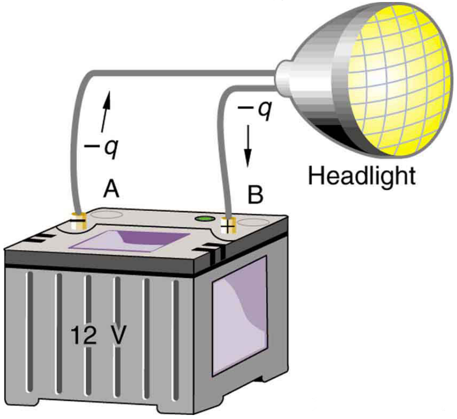
$$\begin{array}{rcl}\Delta\text{PE}_{\text{car}} & = & (60{,}000\text{ C})(12.0\text{ V})\\ & = & 7.20 \times 10^{5}\text{ J.}\end{array}$$

Discussion

While voltage and energy are related, they are not the same thing. The voltages of the batteries are identical, but the energy supplied by each is quite different. Note also that as a battery is discharged, some of its energy is used internally and its terminal voltage drops, such as when headlights dim because of a low car battery. The energy supplied by the battery is still calculated as in this example, but not all of the energy is available for external use.

Note that the energies calculated in the previous example are absolute values. The change in potential energy for the battery is negative, since it loses energy. These batteries, like many electrical systems, actually move negative charge—electrons in particular. The batteries repel electrons from their negative terminals (A) through whatever circuitry is involved and attract them to their positive terminals (B) as shown in [link]. The change in potential is $\Delta V = V_{\text{B}} - V_{\text{A}} = +12\text{ V}$ and the charge q is negative, so that

$\Delta\text{PE} = q\Delta V$ is negative, meaning the potential energy of the battery has decreased when q has moved from A to B.



A battery moves negative charge from its negative terminal through a headlight to its positive terminal. Appropriate combinations of chemicals in the battery separate charges so that the negative terminal has an excess of negative charge, which is repelled by it and attracted to the excess positive charge on the other terminal. In terms of potential, the positive terminal is at a higher voltage than the negative. Inside the battery, both positive and negative charges move.

Example:

How Many Electrons Move through a Headlight Each Second?

When a 12.0 V car battery runs a single 30.0 W headlight, how many electrons pass through it each second?

Strategy

To find the number of electrons, we must first find the charge that moved in 1.00 s. The charge moved is related to voltage and energy through the equation $\Delta\text{PE} = q\Delta V$. A 30.0 W lamp uses 30.0 joules per second. Since the battery loses energy, we have $\Delta\text{PE} = –30.0\text{ J}$ and, since the electrons are going from the negative terminal to the positive, we see that $\Delta V = +12.0\text{ V}$.

Solution

To find the charge q moved, we solve the equation $\Delta\text{PE} = q\Delta V$:

Equation:

$$q = \frac{\Delta\text{PE}}{\Delta V}.$$

Entering the values for ΔPE and ΔV, we get

Equation:

$$q = \frac{–30.0\text{ J}}{+12.0\text{ V}} = \frac{–30.0\text{ J}}{+12.0\text{ J/C}} = –2.50\text{ C}.$$

The number of electrons n_e is the total charge divided by the charge per electron. That is,

Equation:

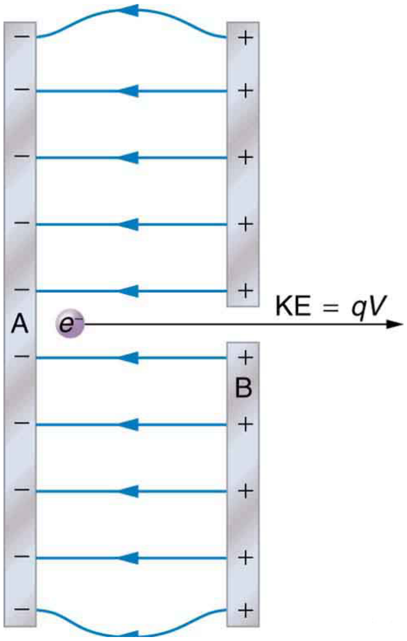
$$\text{n}_\text{e} = \frac{–2.50\text{ C}}{–1.60\times10^{–19}\text{ C/e}^{–}} = 1.56\times10^{19}\text{ electrons.}$$

Discussion

This is a very large number. It is no wonder that we do not ordinarily observe individual electrons with so many being present in ordinary systems. In fact, electricity had been in use for many decades before it was determined that the moving charges in many circumstances were negative. Positive charge moving in the opposite direction of negative charge often produces identical effects; this makes it difficult to determine which is moving or whether both are moving.

The Electron Volt

The energy per electron is very small in macroscopic situations like that in the previous example—a tiny fraction of a joule. But on a submicroscopic scale, such energy per particle (electron, proton, or ion) can be of great importance. For example, even a tiny fraction of a joule can be great enough for these particles to destroy organic molecules and harm living tissue. The particle may do its damage by direct collision, or it may create harmful x rays, which can also inflict damage. It is useful to have an energy unit related to submicroscopic effects. [link] shows a situation related to the definition of such an energy unit. An electron is accelerated between two charged metal plates as it might be in an old-model television tube or oscilloscope. The electron is given kinetic energy that is later converted to another form—light in the television tube, for example. (Note that downhill for the electron is uphill for a positive charge.) Since energy is related to voltage by $\Delta \text{PE} = q\Delta V$, we can think of the joule as a coulomb-volt.



A typical electron gun accelerates electrons using a potential difference between two metal plates. The energy of the electron in electron volts is numerically the same as the voltage between the plates. For example, a 5000 V potential difference produces 5000 eV electrons.

On the submicroscopic scale, it is more convenient to define an energy unit called the **electron volt** (eV), which is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,

Equation:

$$
\begin{aligned}
1\ \text{eV} &= \left(1.60\times 10^{-19}\ \text{C}\right)(1\ \text{V}) = \left(1.60{\times}10^{-19}\ \text{C}\right)(1\ \text{J/C}) \\
&= 1.60\times 10^{-19}\ \text{J}.
\end{aligned}
$$

Note:

Electron Volt

On the submicroscopic scale, it is more convenient to define an energy unit called the electron volt (eV), which is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,

Equation:

$$
\begin{aligned}
1\ \text{eV} &= \left(1.60\times 10^{-19}\ \text{C}\right)(1\ \text{V}) = \left(1.60{\times}10^{-19}\ \text{C}\right)(1\ \text{J/C}) \\
&= 1.60\times 10^{-19}\ \text{J}.
\end{aligned}
$$

An electron accelerated through a potential difference of 1 V is given an energy of 1 eV. It follows that an electron accelerated through 50 V is given 50 eV. A potential difference of 100,000 V (100 kV) will give an electron an energy of 100,000 eV (100 keV), and so on. Similarly, an ion with a double positive charge accelerated through 100 V will be given 200 eV of energy. These simple relationships between accelerating voltage and particle charges make the electron volt a simple and convenient energy unit in such circumstances.

Note:

Connections: Energy Units

The electron volt (eV) is the most common energy unit for submicroscopic processes. This will be particularly noticeable in the chapters on modern physics. Energy is so important to so many subjects that there is a tendency to define a special energy unit for each major topic. There are, for example, calories for food energy, kilowatt-hours for electrical energy, and therms for natural gas energy.

The electron volt is commonly employed in submicroscopic processes—chemical valence energies and molecular and nuclear binding energies are among the quantities often expressed in electron volts. For example, about 5 eV of energy is required to break up certain organic molecules. If a proton is accelerated from rest through a potential difference of 30 kV, it is given an energy of 30 keV (30,000 eV) and it can break up as many as 6000 of these molecules ($30{,}000\text{ eV} \div 5\text{ eV per molecule} = 6000\text{ molecules}$). Nuclear decay energies are on the order of 1 MeV (1,000,000 eV) per event and can, thus, produce significant biological damage.

Conservation of Energy

The total energy of a system is conserved if there is no net addition (or subtraction) of work or heat transfer. For conservative forces, such as the electrostatic force, conservation of energy states that mechanical energy is a constant.

Mechanical energy is the sum of the kinetic energy and potential energy of a system; that is, $\text{KE} + \text{PE} = \text{constant}$. A loss of PE of a charged particle becomes an increase in its KE. Here PE is the electric potential energy. Conservation of energy is stated in equation form as

Equation:

$$\text{KE} + \text{PE} = \text{constant}$$

or

Equation:

$$\text{KE}_\text{i} + \text{PE}_\text{i} = \text{KE}_\text{f} + \text{PE}_\text{f},$$

where i and f stand for initial and final conditions. As we have found many times before, considering energy can give us insights and facilitate problem solving.

Example:

Electrical Potential Energy Converted to Kinetic Energy

Calculate the final speed of a free electron accelerated from rest through a potential difference of 100 V. (Assume that this numerical value is accurate to three significant figures.)

Strategy

We have a system with only conservative forces. Assuming the electron is accelerated in a vacuum, and neglecting the gravitational force (we will check on this assumption later), all of the electrical potential energy is converted into kinetic energy. We can identify the initial and final forms of energy to be $\mathrm{KE_i} = 0$, $\mathrm{KE_f} = ½\,mv^2$, $\mathrm{PE_i} = qV$, and $\mathrm{PE_f} = 0$.

Solution

Conservation of energy states that

Equation:

$$\mathrm{KE_i} + \mathrm{PE_i} = \mathrm{KE_f} + \mathrm{PE_f}.$$

Entering the forms identified above, we obtain

Equation:

$$qV = \frac{mv^2}{2}.$$

We solve this for v:

Equation:

$$v = \sqrt{\frac{2\mathrm{qV}}{m}}.$$

Entering values for q, V, and m gives

Equation:

$$\begin{aligned} v &= \sqrt{\frac{2\left(-1.60\times10^{-19}\text{ C}\right)\left(-100\text{ J/C}\right)}{9.11\times10^{-31}\text{ kg}}} \\ &= 5.93 \times 10^{6}\text{ m/s.} \end{aligned}$$

Discussion

Note that both the charge and the initial voltage are negative, as in [link]. From the discussions in Electric Charge and Electric Field, we know that electrostatic forces on small particles are generally very large compared with the gravitational force. The large final speed confirms that the gravitational force is indeed negligible here. The large speed also indicates how easy it is to accelerate electrons with small voltages because of their very small mass. Voltages much higher than the 100 V in this problem are typically used in electron guns. Those higher voltages produce electron speeds so great that relativistic effects must be taken into account. That is why a low voltage is considered (accurately) in this example.

Section Summary

- Electric potential is potential energy per unit charge.
- The potential difference between points A and B, $V_\text{B}-V_\text{A}$, defined to be the change in potential energy of a charge q moved from A to B, is equal to the change in potential energy divided by the charge, Potential difference is commonly called voltage, represented by the symbol ΔV.
  Equation:

$$\Delta V = \frac{\Delta\text{PE}}{q}\text{ and }\Delta\text{PE} = q\Delta V.$$

- An electron volt is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,
  Equation:

$$\begin{aligned} 1\text{ eV} &= \left(1.60 \times 10^{-19}\text{ C}\right)\left(1\text{ V}\right) = \left(1.60\times10^{-19}\text{ C}\right)\left(1\text{ J/C}\right) \\ &= 1.60 \times 10^{-19}\text{ J.} \end{aligned}$$

- Mechanical energy is the sum of the kinetic energy and potential energy of a system, that is, $\text{KE} + \text{PE}$. This sum is a constant.

Conceptual Questions

Exercise:

Problem:

Voltage is the common word for potential difference. Which term is more descriptive, voltage or potential difference?

Exercise:

Problem:

If the voltage between two points is zero, can a test charge be moved between them with zero net work being done? Can this necessarily be done without exerting a force? Explain.

Exercise:

Problem:

What is the relationship between voltage and energy? More precisely, what is the relationship between potential difference and electric potential energy?

Exercise:

Problem: Voltages are always measured between two points. Why?

Exercise:

Problem:

How are units of volts and electron volts related? How do they differ?

Problems & Exercises

Exercise:

Problem:

Find the ratio of speeds of an electron and a negative hydrogen ion (one having an extra electron) accelerated through the same voltage, assuming non-relativistic final speeds. Take the mass of the hydrogen ion to be 1.67×10^{-27} kg.

Solution:

42.8

Exercise:

Problem:

An evacuated tube uses an accelerating voltage of 40 kV to accelerate electrons to hit a copper plate and produce x rays. Non-relativistically, what would be the maximum speed of these electrons?

Exercise:

Problem:

A bare helium nucleus has two positive charges and a mass of 6.64×10^{-27} kg. (a) Calculate its kinetic energy in joules at 2.00% of the speed of light. (b) What is this in electron volts? (c) What voltage would be needed to obtain this energy?

Exercise:

Problem: Integrated Concepts

Singly charged gas ions are accelerated from rest through a voltage of 13.0 V. At what temperature will the average kinetic energy of gas molecules be the same as that given these ions?

Solution:

1.00×10^{5} K

Exercise:

Problem: **Integrated Concepts**

The temperature near the center of the Sun is thought to be 15 million degrees Celsius $\left(1.5\times10^{7}\ ^{\text{o}}\text{C}\right)$. Through what voltage must a singly charged ion be accelerated to have the same energy as the average kinetic energy of ions at this temperature?

Exercise:

Problem: **Integrated Concepts**

(a) What is the average power output of a heart defibrillator that dissipates 400 J of energy in 10.0 ms? (b) Considering the high-power output, why doesn't the defibrillator produce serious burns?

Solution:

(a) $4\times10^{4}\text{ W}$

(b) A defibrillator does not cause serious burns because the skin conducts electricity well at high voltages, like those used in defibrillators. The gel used aids in the transfer of energy to the body, and the skin doesn't absorb the energy, but rather lets it pass through to the heart.

Exercise:

Problem: **Integrated Concepts**

A lightning bolt strikes a tree, moving 20.0 C of charge through a potential difference of $1.00\times10^{2}\text{ MV}$. (a) What energy was dissipated? (b) What mass of water could be raised from 15^{o}C to the boiling point and then boiled by this energy? (c) Discuss the damage that could be caused to the tree by the expansion of the boiling steam.

Exercise:

Problem: **Integrated Concepts**

A 12.0 V battery-operated bottle warmer heats 50.0 g of glass, 2.50×10^2 g of baby formula, and 2.00×10^2 g of aluminum from 20.0°C to 90.0°C. (a) How much charge is moved by the battery? (b) How many electrons per second flow if it takes 5.00 min to warm the formula? (Hint: Assume that the specific heat of baby formula is about the same as the specific heat of water.)

Solution:

(a) $7.40 \times 10^3\ \text{C}$

(b) 1.54×10^{20} electrons per second

Exercise:

Problem: **Integrated Concepts**

A battery-operated car utilizes a 12.0 V system. Find the charge the batteries must be able to move in order to accelerate the 750 kg car from rest to 25.0 m/s, make it climb a 2.00×10^2 m high hill, and then cause it to travel at a constant 25.0 m/s by exerting a 5.00×10^2 N force for an hour.

Solution:

$3.89 \times 10^6\ \text{C}$

Exercise:

Problem: **Integrated Concepts**

Fusion probability is greatly enhanced when appropriate nuclei are brought close together, but mutual Coulomb repulsion must be overcome. This can be done using the kinetic energy of high-temperature gas ions or by accelerating the nuclei toward one another.

(a) Calculate the potential energy of two singly charged nuclei separated by $1.00\times10^{-12}\ \text{m}$ by finding the voltage of one at that distance and multiplying by the charge of the other. (b) At what temperature will atoms of a gas have an average kinetic energy equal to this needed electrical potential energy?

Exercise:

Problem: Unreasonable Results

(a) Find the voltage near a 10.0 cm diameter metal sphere that has 8.00 C of excess positive charge on it. (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Solution:

(a) $1.44 \times 10^{12}\ \text{V}$

(b) This voltage is very high. A 10.0 cm diameter sphere could never maintain this voltage; it would discharge.

(c) An 8.00 C charge is more charge than can reasonably be accumulated on a sphere of that size.

Exercise:

Problem: Construct Your Own Problem

Consider a battery used to supply energy to a cellular phone. Construct a problem in which you determine the energy that must be supplied by the battery, and then calculate the amount of charge it must be able to move in order to supply this energy. Among the things to be considered are the energy needs and battery voltage. You may need to look ahead to interpret manufacturer’s battery ratings in ampere-hours as energy in joules.

Glossary

electric potential
potential energy per unit charge

potential difference (or voltage)
change in potential energy of a charge moved from one point to another, divided by the charge; units of potential difference are joules per coulomb, known as volt

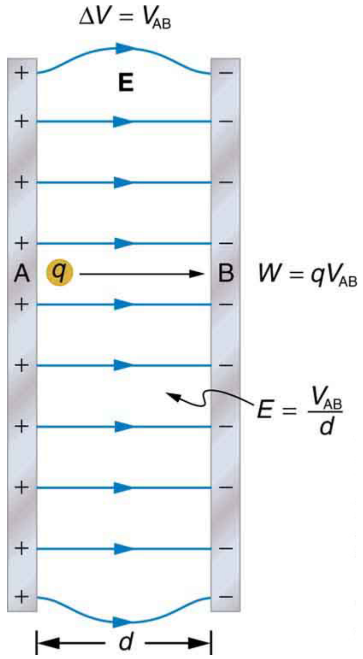
electron volt
the energy given to a fundamental charge accelerated through a potential difference of one volt

mechanical energy
sum of the kinetic energy and potential energy of a system; this sum is a constant

Electric Potential in a Uniform Electric Field

- Describe the relationship between voltage and electric field.
- Derive an expression for the electric potential and electric field.
- Calculate electric field strength given distance and voltage.

In the previous section, we explored the relationship between voltage and energy. In this section, we will explore the relationship between voltage and electric field. For example, a uniform electric field $\mathbf{E}$ is produced by placing a potential difference (or voltage) ΔV across two parallel metal plates, labeled A and B. (See [link].) Examining this will tell us what voltage is needed to produce a certain electric field strength; it will also reveal a more fundamental relationship between electric potential and electric field. From a physicist’s point of view, either ΔV or $\mathbf{E}$ can be used to describe any charge distribution. ΔV is most closely tied to energy, whereas $\mathbf{E}$ is most closely related to force. ΔV is a **scalar** quantity and has no direction, while $\mathbf{E}$ is a **vector** quantity, having both magnitude and direction. (Note that the magnitude of the electric field strength, a scalar quantity, is represented by E below.) The relationship between ΔV and $\mathbf{E}$ is revealed by calculating the work done by the force in moving a charge from point A to point B. But, as noted in Electric Potential Energy: Potential Difference, this is complex for arbitrary charge distributions, requiring calculus. We therefore look at a uniform electric field as an interesting special case.



The relationship between V and E for parallel conducting plates is $E = V/d$. (Note that $\Delta V = V_{\mathrm{AB}}$ in magnitude. For a charge that is moved from plate A at higher potential to plate B at lower potential, a minus sign needs to be included as follows: $-\Delta V = V_{\mathrm{A}} - V_{\mathrm{B}} = V_{\mathrm{AB}}$. See the text for details.)

The work done by the electric field in [link] to move a positive charge q from A, the positive plate, higher potential, to B, the negative plate, lower potential, is

Equation:

$$W = -\Delta\text{PE} = -q\Delta V.$$

The potential difference between points A and B is

Equation:

$$-\Delta\ V = -(V_{\text{B}} - V_{\text{A}}) = V_{\text{A}} - V_{\text{B}} = V_{\text{AB}}.$$

Entering this into the expression for work yields

Equation:

$$W = qV_{\text{AB}}.$$

Work is $W = Fd\cos\theta$; here $\cos\theta = 1$, since the path is parallel to the field, and so $W = Fd$. Since $F = qE$, we see that $W = qEd$. Substituting this expression for work into the previous equation gives

Equation:

$$qEd = qV_{\text{AB}}.$$

The charge cancels, and so the voltage between points A and B is seen to be

Equation:

$$\left.\begin{array}{c} V_{\text{AB}} = Ed \\ E = \frac{V_{\text{AB}}}{d} \end{array}\right\}(\text{uniform } E \text{ - field only}),$$

where d is the distance from A to B, or the distance between the plates in [link]. Note that the above equation implies the units for electric field are volts per meter. We already know the units for electric field are newtons per coulomb; thus the following relation among units is valid:

Equation:

$$1\ \text{N/C} = 1\ \text{V/m}.$$

Note:
Voltage between Points A and B
Equation:

$$\left.\begin{array}{r}V_{\text{AB}}=Ed\\ E=\frac{V_{\text{AB}}}{d}\end{array}\right\}(\text{uniform }E\text{ - field only}),$$

where d is the distance from A to B, or the distance between the plates.

Example:
What Is the Highest Voltage Possible between Two Plates?
Dry air will support a maximum electric field strength of about $3.0\times10^{6}\text{ V/m}$. Above that value, the field creates enough ionization in the air to make the air a conductor. This allows a discharge or spark that reduces the field. What, then, is the maximum voltage between two parallel conducting plates separated by 2.5 cm of dry air?
Strategy
We are given the maximum electric field E between the plates and the distance d between them. The equation $V_{\text{AB}}=Ed$ can thus be used to calculate the maximum voltage.
Solution
The potential difference or voltage between the plates is
Equation:

$$\text{V}_{\text{AB}}=Ed.$$

Entering the given values for E and d gives
Equation:

$$V_{\text{AB}}=(3.0\times10^{6}\text{ V/m})(0.025\text{ m})=7.5\times10^{4}\text{ V}$$

or
Equation:

$$V_{\text{AB}}=75\text{ kV}.$$

(The answer is quoted to only two digits, since the maximum field strength is approximate.)

Discussion

One of the implications of this result is that it takes about 75 kV to make a spark jump across a 2.5 cm (1 in.) gap, or 150 kV for a 5 cm spark. This limits the voltages that can exist between conductors, perhaps on a power transmission line. A smaller voltage will cause a spark if there are points on the surface, since points create greater fields than smooth surfaces. Humid air breaks down at a lower field strength, meaning that a smaller voltage will make a spark jump through humid air. The largest voltages can be built up, say with static electricity, on dry days.

scintillation counter
high voltage switch
scintillation counter

A spark chamber is used to trace the paths of high-energy particles. Ionization created by the particles as they pass through the gas between the plates allows a spark to jump. The sparks are perpendicular to the plates, following electric field lines between them. The potential difference between adjacent plates is not high enough to cause sparks without the ionization produced by particles from accelerator experiments (or cosmic rays). (credit: Daderot, Wikimedia Commons)

Example:

Field and Force inside an Electron Gun

(a) An electron gun has parallel plates separated by 4.00 cm and gives electrons 25.0 keV of energy. What is the electric field strength between the plates? (b) What force would this field exert on a piece of plastic with a $0.500\ \mu\text{C}$ charge that gets between the plates?

Strategy

Since the voltage and plate separation are given, the electric field strength can be calculated directly from the expression $E = \frac{V_{\text{AB}}}{d}$. Once the electric field strength is known, the force on a charge is found using $\mathbf{F} = q\,\mathbf{E}$. Since the electric field is in only one direction, we can write this equation in terms of the magnitudes, $F = q\,E$.

Solution for (a)

The expression for the magnitude of the electric field between two uniform metal plates is

Equation:

$$E = \frac{V_{\text{AB}}}{d}.$$

Since the electron is a single charge and is given 25.0 keV of energy, the potential difference must be 25.0 kV. Entering this value for V_{AB} and the plate separation of 0.0400 m, we obtain

Equation:

$$E = \frac{25.0\text{ kV}}{0.0400\text{ m}} = 6.25 \times 10^5\text{ V/m}.$$

Solution for (b)

The magnitude of the force on a charge in an electric field is obtained from the equation

Equation:

$$F = qE.$$

Substituting known values gives

Equation:

$$F = (0.500 \times 10^{-6}\text{ C})(6.25 \times 10^5\text{ V/m}) = 0.313\text{ N}.$$

Discussion

Note that the units are newtons, since $1\text{ V/m} = 1\text{ N/C}$. The force on the charge is the same no matter where the charge is located between the plates. This is because the electric field is uniform between the plates.

In more general situations, regardless of whether the electric field is uniform, it points in the direction of decreasing potential, because the force on a positive charge is in the direction of $\mathbf{E}$ and also in the direction of lower potential V. Furthermore, the magnitude of $\mathbf{E}$ equals the rate of decrease of V with distance. The faster V decreases over distance, the greater the electric field. In equation form, the general relationship between voltage and electric field is

Equation:

$$E = -\frac{\Delta V}{\Delta s},$$

where Δs is the distance over which the change in potential, ΔV, takes place. The minus sign tells us that $\mathbf{E}$ points in the direction of decreasing potential. The electric field is said to be the *gradient* (as in grade or slope) of the electric potential.

Note:

Relationship between Voltage and Electric Field

In equation form, the general relationship between voltage and electric field is

Equation:

$$E = -\frac{\Delta V}{\Delta s},$$

where Δs is the distance over which the change in potential, ΔV, takes place. The minus sign tells us that $\mathbf{E}$ points in the direction of decreasing potential. The electric field is said to be the *gradient* (as in grade or slope) of the electric potential.

For continually changing potentials, ΔV and Δs become infinitesimals and differential calculus must be employed to determine the electric field.

Section Summary

- The voltage between points A and B is
  Equation:

$$\left.\begin{array}{l} V_{\mathrm{AB}} = Ed \\ E = \frac{V_{\mathrm{AB}}}{d} \end{array}\right\}(\text{uniform } E \text{ - field only}),$$

where d is the distance from A to B, or the distance between the plates.

- In equation form, the general relationship between voltage and electric field is

  Equation:

  $$E = -\frac{\Delta V}{\Delta s},$$

  where Δs is the distance over which the change in potential, ΔV, takes place. The minus sign tells us that $\mathbf{E}$ points in the direction of decreasing potential.) The electric field is said to be the *gradient* (as in grade or slope) of the electric potential.

Conceptual Questions

Exercise:

Problem:

Discuss how potential difference and electric field strength are related. Give an example.

Exercise:

Problem:

What is the strength of the electric field in a region where the electric potential is constant?

Exercise:

Problem:

Will a negative charge, initially at rest, move toward higher or lower potential? Explain why.

Problems & Exercises

Exercise:

Problem:

Show that units of V/m and N/C for electric field strength are indeed equivalent.

Exercise:

Problem:

What is the strength of the electric field between two parallel conducting plates separated by 1.00 cm and having a potential difference (voltage) between them of $1.50 \times 10^4 \text{ V}$?

Exercise:

Problem:

The electric field strength between two parallel conducting plates separated by 4.00 cm is $7.50 \times 10^4 \text{ V/m}$. (a) What is the potential difference between the plates? (b) The plate with the lowest potential is taken to be at zero volts. What is the potential 1.00 cm from that plate (and 3.00 cm from the other)?

Solution:

(a) 3.00 kV

(b) 750 V

Exercise:

Problem:

How far apart are two conducting plates that have an electric field strength of $4.50 \times 10^3 \text{ V/m}$ between them, if their potential difference is 15.0 kV?

Exercise:

Problem:

(a) Will the electric field strength between two parallel conducting plates exceed the breakdown strength for air $(3.0 \times 10^6\ \text{V/m})$ if the plates are separated by 2.00 mm and a potential difference of $5.0 \times 10^3\ \text{V}$ is applied? (b) How close together can the plates be with this applied voltage?

Solution:

(a) No. The electric field strength between the plates is $2.5 \times 10^6\ \text{V/m}$, which is lower than the breakdown strength for air ($3.0 \times 10^6\ \text{V/m}$).

(b) 1.7 mm

Exercise:

Problem:

The voltage across a membrane forming a cell wall is 80.0 mV and the membrane is 9.00 nm thick. What is the electric field strength? (The value is surprisingly large, but correct. Membranes are discussed in Capacitors and Dielectrics and Nerve Conduction—Electrocardiograms.) You may assume a uniform electric field.

Exercise:

Problem:

Membrane walls of living cells have surprisingly large electric fields across them due to separation of ions. (Membranes are discussed in some detail in Nerve Conduction—Electrocardiograms.) What is the voltage across an 8.00 nm–thick membrane if the electric field strength across it is 5.50 MV/m? You may assume a uniform electric field.

Solution:

44.0 mV

Exercise:

Problem:

Two parallel conducting plates are separated by 10.0 cm, and one of them is taken to be at zero volts. (a) What is the electric field strength between them, if the potential 8.00 cm from the zero volt plate (and 2.00 cm from the other) is 450 V? (b) What is the voltage between the plates?

Exercise:

Problem:

Find the maximum potential difference between two parallel conducting plates separated by 0.500 cm of air, given the maximum sustainable electric field strength in air to be $3.0 \times 10^6 \text{ V/m}$.

Solution:

15 kV

Exercise:

Problem:

A doubly charged ion is accelerated to an energy of 32.0 keV by the electric field between two parallel conducting plates separated by 2.00 cm. What is the electric field strength between the plates?

Exercise:

Problem:

An electron is to be accelerated in a uniform electric field having a strength of $2.00 \times 10^6 \text{ V/m}$. (a) What energy in keV is given to the electron if it is accelerated through 0.400 m? (b) Over what distance would it have to be accelerated to increase its energy by 50.0 GeV?

Solution:

(a) 800 KeV

(b) 25.0 km

Glossary

scalar	physical quantity with magnitude but no direction
vector	physical quantity with both magnitude and direction

Electrical Potential Due to a Point Charge

- Explain point charges and express the equation for electric potential of a point charge.
- Distinguish between electric potential and electric field.
- Determine the electric potential of a point charge given charge and distance.

Point charges, such as electrons, are among the fundamental building blocks of matter. Furthermore, spherical charge distributions (like on a metal sphere) create external electric fields exactly like a point charge. The electric potential due to a point charge is, thus, a case we need to consider. Using calculus to find the work needed to move a test charge q from a large distance away to a distance of r from a point charge Q, and noting the connection between work and potential $(W=-q\Delta V)$, it can be shown that the *electric potential V of a point charge* is

Equation:

$$V=\frac{kQ}{r}\;\text{(Point Charge)},$$

where *k* is a constant equal to $9.0\times 10^{9}\ \text{N}\cdot\text{m}^2/\text{C}^2$.

Note:

Electric Potential V of a Point Charge

The electric potential V of a point charge is given by

Equation:

$$V=\frac{kQ}{r}\;\text{(Point Charge)}.$$

The potential at infinity is chosen to be zero. Thus V for a point charge decreases with distance, whereas $\mathbf{E}$ for a point charge decreases with distance squared:

Equation:

$$E = \frac{F}{q} = \frac{kQ}{r^2}.$$

Recall that the electric potential V is a scalar and has no direction, whereas the electric field $\mathbf{E}$ is a vector. To find the voltage due to a combination of point charges, you add the individual voltages as numbers. To find the total electric field, you must add the individual fields as *vectors*, taking magnitude and direction into account. This is consistent with the fact that V is closely associated with energy, a scalar, whereas $\mathbf{E}$ is closely associated with force, a vector.

Example:

What Voltage Is Produced by a Small Charge on a Metal Sphere?

Charges in static electricity are typically in the nanocoulomb (nC) to microcoulomb (μC) range. What is the voltage 5.00 cm away from the center of a 1-cm diameter metal sphere that has a $-3.00\ \text{nC}$ static charge?

Strategy

As we have discussed in Electric Charge and Electric Field, charge on a metal sphere spreads out uniformly and produces a field like that of a point charge located at its center. Thus we can find the voltage using the equation $V = \text{kQ}/r$.

Solution

Entering known values into the expression for the potential of a point charge, we obtain

Equation:

$$\begin{aligned} V &= k\frac{Q}{r} \\ &= \left(8.99\times10^{9}\ \text{N}\cdot\text{m}^2/\text{C}^2\right)\left(\frac{-3.00\times10^{-9}\ \text{C}}{5.00\times10^{-2}\ \text{m}}\right) \\ &= -539\ \text{V}. \end{aligned}$$

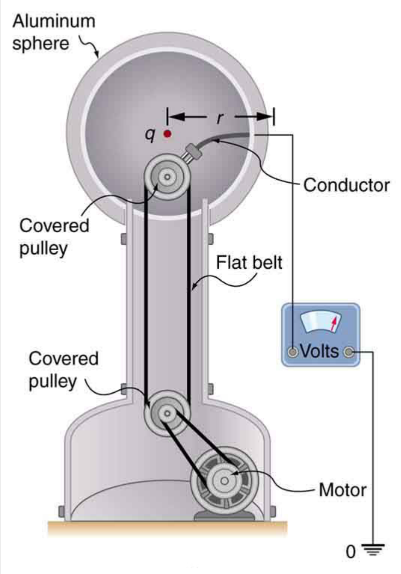
Discussion

The negative value for voltage means a positive charge would be attracted from a larger distance, since the potential is lower (more negative) than at larger distances. Conversely, a negative charge would be repelled, as expected.

Example:

What Is the Excess Charge on a Van de Graaff Generator

A demonstration Van de Graaff generator has a 25.0 cm diameter metal sphere that produces a voltage of 100 kV near its surface. (See [link].) What excess charge resides on the sphere? (Assume that each numerical value here is shown with three significant figures.)



The voltage of this demonstration Van de Graaff generator is measured between the charged sphere and ground. Earth’s potential is taken to be zero as a reference. The potential of the charged conducting sphere is the same as that of an equal point charge at its center.

Strategy

The potential on the surface will be the same as that of a point charge at the center of the sphere, 12.5 cm away. (The radius of the sphere is 12.5 cm.) We can thus determine the excess charge using the equation

Equation:

$$V=\frac{\text{kQ}}{r}.$$

Solution

Solving for Q and entering known values gives

Equation:

$$\begin{array}{lll}Q & = & \frac{\text{rV}}{k}\\ & = & \frac{\left(0\text{.}\text{125 m}\right)\left(\text{100}\times {\text{10}}^{3}\text{ V}\right)}{8\text{.}\text{99}\times {\text{10}}^{9}\text{ N}\cdot {\text{m}}^{2}{\text{/C}}^{2}}\\ & = & 1\text{.}\text{39}\times {\text{10}}^{-6}\text{ C}=1\text{.}\text{39 µC.}\end{array}$$

Discussion

This is a relatively small charge, but it produces a rather large voltage. We have another indication here that it is difficult to store isolated charges.

The voltages in both of these examples could be measured with a meter that compares the measured potential with ground potential. Ground potential is often taken to be zero (instead of taking the potential at infinity to be zero). It is the potential difference between two points that is of importance, and very often there is a tacit assumption that some reference point, such as Earth or a very distant point, is at zero potential. As noted in Electric Potential Energy: Potential Difference, this is analogous to taking sea level as $h = 0$ when considering gravitational potential energy, $\mathrm{PE_g} = \mathrm{mgh}$.

Section Summary

- Electric potential of a point charge is $V = \mathrm{kQ}/r$.
- Electric potential is a scalar, and electric field is a vector. Addition of voltages as numbers gives the voltage due to a combination of point charges, whereas addition of individual fields as vectors gives the total electric field.

Conceptual Questions

Exercise:

Problem:

In what region of space is the potential due to a uniformly charged sphere the same as that of a point charge? In what region does it differ from that of a point charge?

Exercise:

Problem:

Can the potential of a non-uniformly charged sphere be the same as that of a point charge? Explain.

Problems & Exercises

Exercise:

Problem:

A 0.500 cm diameter plastic sphere, used in a static electricity demonstration, has a uniformly distributed 40.0 pC charge on its surface. What is the potential near its surface?

Solution:

144 V

Exercise:

Problem:

What is the potential $0.530 \times 10^{-10}\ \text{m}$ from a proton (the average distance between the proton and electron in a hydrogen atom)?

Exercise:

Problem:

(a) A sphere has a surface uniformly charged with 1.00 C. At what distance from its center is the potential 5.00 MV? (b) What does your answer imply about the practical aspect of isolating such a large charge?

Solution:

(a) 1.80 km

(b) A charge of 1 C is a very large amount of charge; a sphere of radius 1.80 km is not practical.

Exercise:

Problem:

How far from a $1.00\ \mu\text{C}$ point charge will the potential be 100 V? At what distance will it be $2.00 \times 10^{2}\ \text{V}$?

Exercise:

Problem:

What are the sign and magnitude of a point charge that produces a potential of -2.00 V at a distance of 1.00 mm?

Solution:

$-2.22\times10^{-13}\text{ C}$

Exercise:

Problem:

If the potential due to a point charge is $5.00\times10^{2}\text{ V}$ at a distance of 15.0 m, what are the sign and magnitude of the charge?

Exercise:

Problem:

In nuclear fission, a nucleus splits roughly in half. (a) What is the potential $2.00\times10^{-14}\text{ m}$ from a fragment that has 46 protons in it? (b) What is the potential energy in MeV of a similarly charged fragment at this distance?

Solution:

(a) $3.31\times10^{6}\text{ V}$

(b) 152 MeV

Exercise:

Problem:

A research Van de Graaff generator has a 2.00-m-diameter metal sphere with a charge of 5.00 mC on it. (a) What is the potential near its surface? (b) At what distance from its center is the potential 1.00 MV? (c) An oxygen atom with three missing electrons is released near the Van de Graaff generator. What is its energy in MeV at this distance?

Exercise:

Problem:

An electrostatic paint sprayer has a 0.200-m-diameter metal sphere at a potential of 25.0 kV that repels paint droplets onto a grounded object. (a) What charge is on the sphere? (b) What charge must a 0.100-mg drop of paint have to arrive at the object with a speed of 10.0 m/s?

Solution:

(a) $2.78 \times 10^{-7}\ \text{C}$

(b) $2.00 \times 10^{-10}\ \text{C}$

Exercise:

Problem:

In one of the classic nuclear physics experiments at the beginning of the 20th century, an alpha particle was accelerated toward a gold nucleus, and its path was substantially deflected by the Coulomb interaction. If the energy of the doubly charged alpha nucleus was 5.00 MeV, how close to the gold nucleus (79 protons) could it come before being deflected?

Exercise:

Problem:

(a) What is the potential between two points situated 10 cm and 20 cm from a $3.0\ \mu\text{C}$ point charge? (b) To what location should the point at 20 cm be moved to increase this potential difference by a factor of two?

Exercise:

Problem: Unreasonable Results

(a) What is the final speed of an electron accelerated from rest through a voltage of 25.0 MV by a negatively charged Van de Graaff terminal?

(b) What is unreasonable about this result?

(c) Which assumptions are responsible?

Solution:

(a) $2.96 \times 10^{9}\ \mathrm{m/s}$

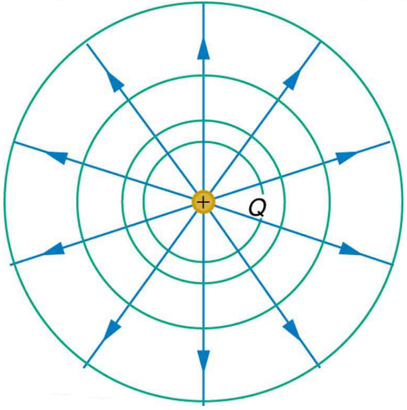
(b) This velocity is far too great. It is faster than the speed of light.

(c) The assumption that the speed of the electron is far less than that of light and that the problem does not require a relativistic treatment produces an answer greater than the speed of light.

Equipotential Lines

- Explain equipotential lines and equipotential surfaces.
- Describe the action of grounding an electrical appliance.
- Compare electric field and equipotential lines.

We can represent electric potentials (voltages) pictorially, just as we drew pictures to illustrate electric fields. Of course, the two are related. Consider [link], which shows an isolated positive point charge and its electric field lines. Electric field lines radiate out from a positive charge and terminate on negative charges. While we use blue arrows to represent the magnitude and direction of the electric field, we use green lines to represent places where the electric potential is constant. These are called **equipotential lines** in two dimensions, or *equipotential surfaces* in three dimensions. The term *equipotential* is also used as a noun, referring to an equipotential line or surface. The potential for a point charge is the same anywhere on an imaginary sphere of radius r surrounding the charge. This is true since the potential for a point charge is given by $V = \mathrm{k}Q/r$ and, thus, has the same value at any point that is a given distance r from the charge. An equipotential sphere is a circle in the two-dimensional view of [link]. Since the electric field lines point radially away from the charge, they are perpendicular to the equipotential lines.



An isolated point charge Q with its electric field lines in blue and equipotential lines in green. The potential is the same along each equipotential line, meaning that no work is required to move a charge anywhere along one of those lines. Work is needed to move a charge from one equipotential line to another. Equipotential lines are perpendicular to electric field lines in every case.

It is important to note that *equipotential lines are always perpendicular to electric field lines*. No work is required to move a charge along an equipotential, since $\Delta V = 0$. Thus the work is

Equation:

$$W = -\Delta \text{ PE} = -q\Delta V = 0.$$

Work is zero if force is perpendicular to motion. Force is in the same direction as $\mathbf{E}$, so that motion along an equipotential must be perpendicular to $\mathbf{E}$. More precisely, work is related to the electric field by

Equation:

$$W = Fd \cos \theta = qEd \cos \theta = 0.$$

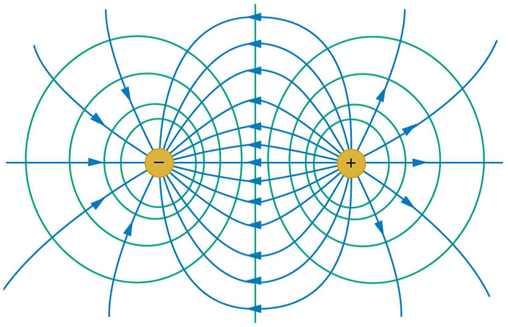
Note that in the above equation, E and F symbolize the magnitudes of the electric field strength and force, respectively. Neither q nor $\mathbf{E}$ nor d is zero, and so $\cos \theta$ must be 0, meaning θ must be 90°. In other words, motion along an equipotential is perpendicular to $\mathbf{E}$.

One of the rules for static electric fields and conductors is that the electric field must be perpendicular to the surface of any conductor. This implies that a *conductor is an equipotential surface in static situations*. There can be no voltage difference across the surface of a conductor, or charges will flow. One of the uses of this fact is that a conductor can be fixed at zero volts by connecting it to the earth with a good conductor—a process called **grounding**. Grounding can be a useful safety tool. For example, grounding the metal case of an electrical appliance ensures that it is at zero volts relative to the earth.

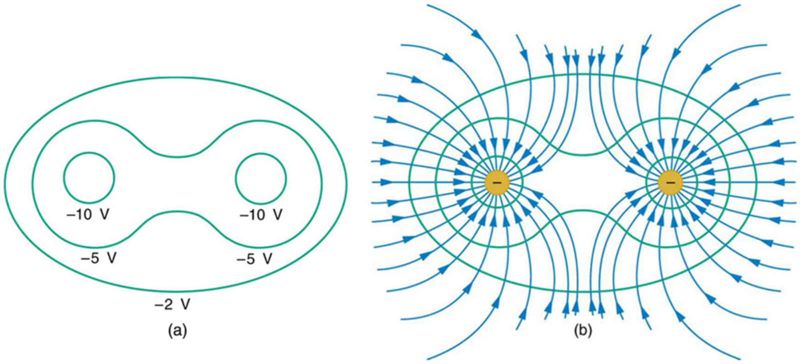
Note:
Grounding
A conductor can be fixed at zero volts by connecting it to the earth with a good conductor—a process called grounding.

Because a conductor is an equipotential, it can replace any equipotential surface. For example, in [link] a charged spherical conductor can replace the point charge, and the electric field and potential surfaces outside of it will be unchanged, confirming the contention that a spherical charge distribution is equivalent to a point charge at its center.

[link] shows the electric field and equipotential lines for two equal and opposite charges. Given the electric field lines, the equipotential lines can be drawn simply by making them perpendicular to the electric field lines. Conversely, given the equipotential lines, as in [link](a), the electric field lines can be drawn by making them perpendicular to the equipotentials, as in [link](b).

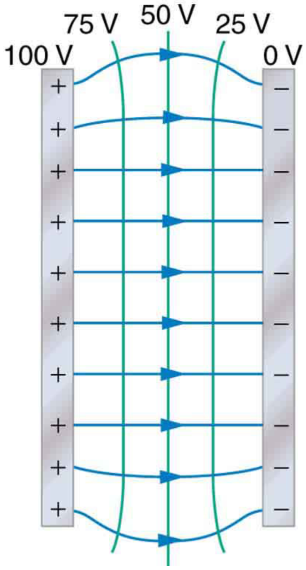


The electric field lines and equipotential lines for two equal but opposite charges. The equipotential lines can be drawn by making them perpendicular to the electric field lines, if those are known. Note that the potential is greatest (most positive) near the positive charge and least (most negative) near the negative charge.



(a) These equipotential lines might be measured with a voltmeter in a laboratory experiment. (b) The corresponding electric field lines are found by drawing them perpendicular to the equipotentials. Note that these fields are consistent with two equal negative charges.

One of the most important cases is that of the familiar parallel conducting plates shown in [link]. Between the plates, the equipotentials are evenly spaced and parallel. The same field could be maintained by placing conducting plates at the equipotential lines at the potentials shown.



The electric field and equipotential lines between two metal plates.

An important application of electric fields and equipotential lines involves the heart. The heart relies on electrical signals to maintain its rhythm. The movement of electrical signals causes the chambers of the heart to contract and relax. When a person has a heart attack, the movement of these electrical signals may be disturbed. An artificial pacemaker and a defibrillator can be used to initiate the rhythm of electrical signals. The equipotential lines around the heart, the thoracic region, and the axis of the heart are useful ways of monitoring the structure and functions of the heart. An electrocardiogram (ECG) measures the small electric signals being generated during the activity of the heart. More about the relationship between electric fields and the heart is discussed in Energy Stored in Capacitors.

Note:
PhET Explorations: Charges and Fields
Move point charges around on the playing field and then view the electric field, voltages, equipotential lines, and more. It's colorful, it's dynamic, it's free.
https://phet.colorado.edu/sims/html/charges-and-fields/latest/charges-and-fields_en.html

Section Summary

- An equipotential line is a line along which the electric potential is constant.
- An equipotential surface is a three-dimensional version of equipotential lines.
- Equipotential lines are always perpendicular to electric field lines.
- The process by which a conductor can be fixed at zero volts by connecting it to the earth with a good conductor is called grounding.

Conceptual Questions

Exercise:

Problem:

What is an equipotential line? What is an equipotential surface?

Exercise:

Problem:

Explain in your own words why equipotential lines and surfaces must be perpendicular to electric field lines.

Exercise:

Problem: Can different equipotential lines cross? Explain.

Problems & Exercises

Exercise:

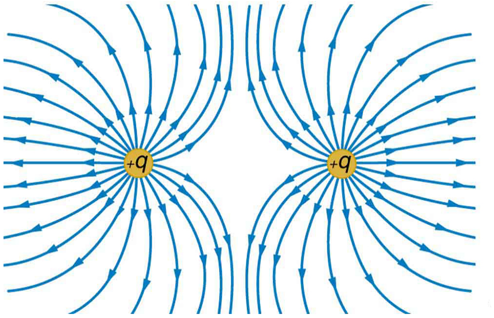
Problem:

(a) Sketch the equipotential lines near a point charge $+q$. Indicate the direction of increasing potential. (b) Do the same for a point charge $-3q$.

Exercise:

Problem:

Sketch the equipotential lines for the two equal positive charges shown in [link]. Indicate the direction of increasing potential.



The electric field near two equal positive charges is directed away from each of the charges.

Exercise:

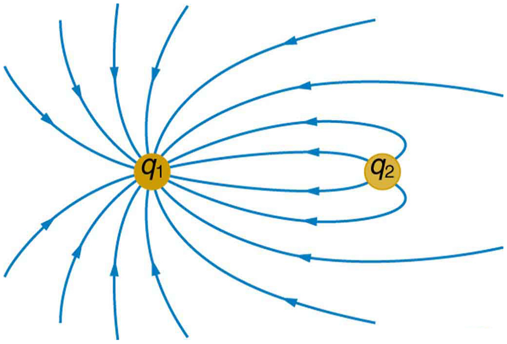
Problem:

[link] shows the electric field lines near two charges q_1 and q_2, the first having a magnitude four times that of the second. Sketch the equipotential lines for these two charges, and indicate the direction of increasing potential.

Exercise:

Problem:

Sketch the equipotential lines a long distance from the charges shown in [link]. Indicate the direction of increasing potential.



The electric field near two charges.

Exercise:

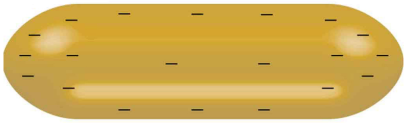
Problem:

Sketch the equipotential lines in the vicinity of two opposite charges, where the negative charge is three times as great in magnitude as the positive. See [link] for a similar situation. Indicate the direction of increasing potential.

Exercise:

Problem:

Sketch the equipotential lines in the vicinity of the negatively charged conductor in [link]. How will these equipotentials look a long distance from the object?

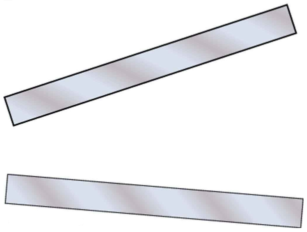


A negatively charged conductor.

Exercise:

Problem:

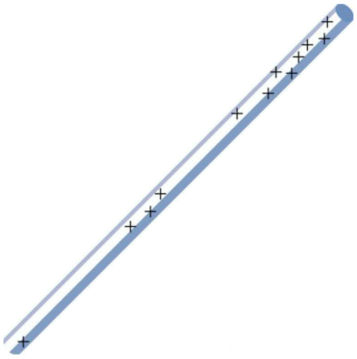
Sketch the equipotential lines surrounding the two conducting plates shown in [link], given the top plate is positive and the bottom plate has an equal amount of negative charge. Be certain to indicate the distribution of charge on the plates. Is the field strongest where the plates are closest? Why should it be?



Exercise:

Problem:

(a) Sketch the electric field lines in the vicinity of the charged insulator in [link]. Note its non-uniform charge distribution. (b) Sketch equipotential lines surrounding the insulator. Indicate the direction of increasing potential.



A charged insulating rod such as might be used in a classroom demonstration.

Exercise:

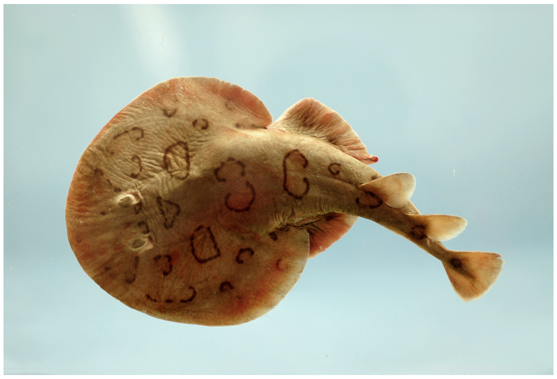
Problem:

The naturally occurring charge on the ground on a fine day out in the open country is –1.00 $\mathrm{nC/m^2}$. (a) What is the electric field relative to ground at a height of 3.00 m? (b) Calculate the electric potential at this height. (c) Sketch electric field and equipotential lines for this scenario.

Exercise:

Problem:

The lesser electric ray (*Narcine bancroftii*) maintains an incredible charge on its head and a charge equal in magnitude but opposite in sign on its tail ([link]). (a) Sketch the equipotential lines surrounding the ray. (b) Sketch the equipotentials when the ray is near a ship with a conducting surface. (c) How could this charge distribution be of use to the ray?



Lesser electric ray (*Narcine bancroftii*) (credit: National Oceanic and Atmospheric Administration, NOAA's Fisheries Collection).

Glossary

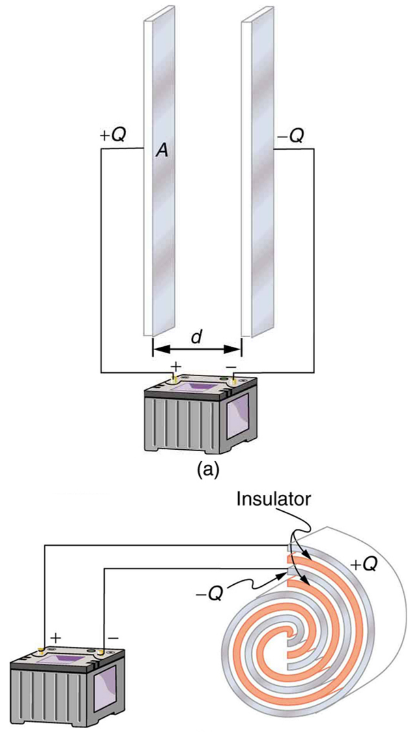
equipotential line	a line along which the electric potential is constant
grounding	fixing a conductor at zero volts by connecting it to the earth or ground

Capacitors and Dielectrics

- Describe the action of a capacitor and define capacitance.
- Explain parallel plate capacitors and their capacitances.
- Discuss the process of increasing the capacitance of a dielectric.
- Determine capacitance given charge and voltage.

A **capacitor** is a device used to store electric charge. Capacitors have applications ranging from filtering static out of radio reception to energy storage in heart defibrillators. Typically, commercial capacitors have two conducting parts close to one another, but not touching, such as those in [link]. (Most of the time an insulator is used between the two plates to provide separation—see the discussion on dielectrics below.) When battery terminals are connected to an initially uncharged capacitor, equal amounts of positive and negative charge, $+Q$ and $-Q$, are separated into its two plates. The capacitor remains neutral overall, but we refer to it as storing a charge Q in this circumstance.

Note:
Capacitor
A capacitor is a device used to store electric charge.



(a)

(b)

Both capacitors shown here were initially uncharged before being connected to a battery. They now have separated charges of $+Q$ and $-Q$ on their two halves. (a) A parallel plate capacitor. (b) A rolled capacitor with an insulating material between its two conducting sheets.

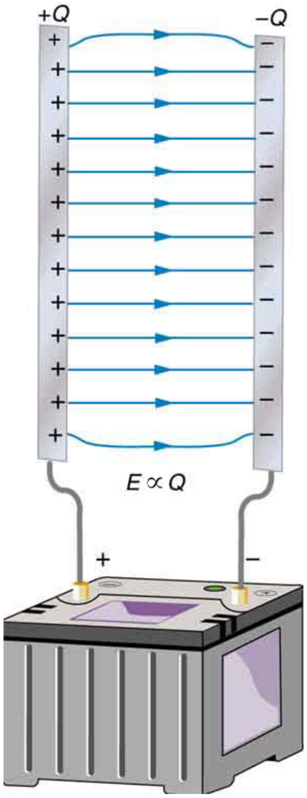
The amount of charge Q a *capacitor* can store depends on two major factors—the voltage applied and the capacitor’s physical characteristics, such as its size.

Note:

The Amount of Charge Q a Capacitor Can Store

The amount of charge Q a *capacitor* can store depends on two major factors—the voltage applied and the capacitor’s physical characteristics, such as its size.

A system composed of two identical, parallel conducting plates separated by a distance, as in [link], is called a **parallel plate capacitor**. It is easy to see the relationship between the voltage and the stored charge for a parallel plate capacitor, as shown in [link]. Each electric field line starts on an individual positive charge and ends on a negative one, so that there will be more field lines if there is more charge. (Drawing a single field line per charge is a convenience, only. We can draw many field lines for each charge, but the total number is proportional to the number of charges.) The electric field strength is, thus, directly proportional to Q.



Electric field lines in this parallel plate capacitor, as always, start on positive charges and end on negative charges. Since the electric field strength is proportional to the density of field lines, it is also proportional to the amount

of charge on
the capacitor.

The field is proportional to the charge:

Equation:

$$E\propto Q,$$

where the symbol $\propto$ means “proportional to.” From the discussion in Electric Potential in a Uniform Electric Field, we know that the voltage across parallel plates is $V=\text{Ed}$. Thus,

Equation:

$$V\propto E.$$

It follows, then, that $V\propto Q$, and conversely,

Equation:

$$Q\propto V.$$

This is true in general: The greater the voltage applied to any capacitor, the greater the charge stored in it.

Different capacitors will store different amounts of charge for the same applied voltage, depending on their physical characteristics. We define their **capacitance** C to be such that the charge Q stored in a capacitor is proportional to C. The charge stored in a capacitor is given by

Equation:

$$Q=\text{CV}.$$

This equation expresses the two major factors affecting the amount of charge stored. Those factors are the physical characteristics of the capacitor,

C, and the voltage, V. Rearranging the equation, we see that *capacitance* C *is the amount of charge stored per volt*, or
Equation:

$$C = \frac{Q}{V}.$$

Note:
Capacitance
Capacitance C is the amount of charge stored per volt, or
Equation:

$$C = \frac{Q}{V}.$$

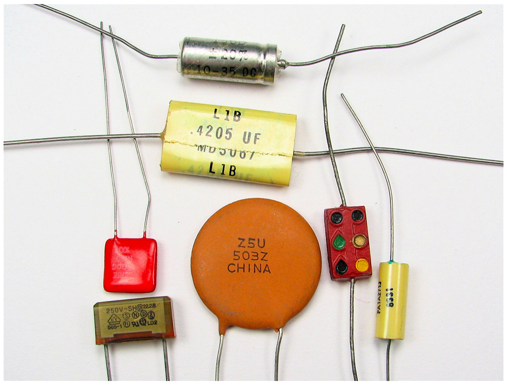
The unit of capacitance is the farad (F), named for Michael Faraday (1791–1867), an English scientist who contributed to the fields of electromagnetism and electrochemistry. Since capacitance is charge per unit voltage, we see that a farad is a coulomb per volt, or
Equation:

$$1\text{ F} = \frac{1\text{ C}}{1\text{ V}}.$$

A 1-farad capacitor would be able to store 1 coulomb (a very large amount of charge) with the application of only 1 volt. One farad is, thus, a very large capacitance. Typical capacitors range from fractions of a picofarad $\left(1\text{ pF} = 10^{-12}\text{ F}\right)$ to millifarads $\left(1\text{ mF} = 10^{-3}\text{ F}\right)$.

[link] shows some common capacitors. Capacitors are primarily made of ceramic, glass, or plastic, depending upon purpose and size. Insulating

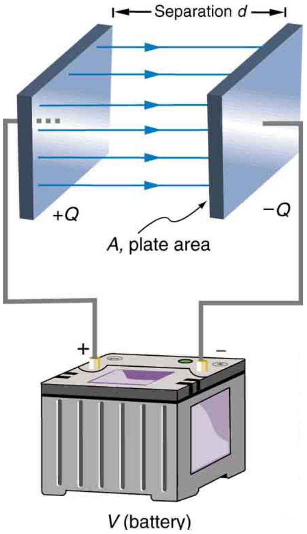
materials, called dielectrics, are commonly used in their construction, as discussed below.



Some typical capacitors. Size and value of capacitance are not necessarily related. (credit: Windell Oskay)

Parallel Plate Capacitor

The parallel plate capacitor shown in [link] has two identical conducting plates, each having a surface area A, separated by a distance d (with no material between the plates). When a voltage V is applied to the capacitor, it stores a charge Q, as shown. We can see how its capacitance depends on A and d by considering the characteristics of the Coulomb force. We know that like charges repel, unlike charges attract, and the force between charges decreases with distance. So it seems quite reasonable that the bigger the plates are, the more charge they can store—because the charges can spread out more. Thus C should be greater for larger A. Similarly, the closer the plates are together, the greater the attraction of the opposite charges on them. So C should be greater for smaller d.



Parallel plate capacitor with plates separated by a distance d. Each plate has an area A.

It can be shown that for a parallel plate capacitor there are only two factors (A and d) that affect its capacitance C. The capacitance of a parallel plate capacitor in equation form is given by

Equation:

$$C = \varepsilon_0 \frac{A}{d}.$$

Note:

Capacitance of a Parallel Plate Capacitor

Equation:

$$C=\varepsilon_0\frac{A}{d}$$

A is the area of one plate in square meters, and d is the distance between the plates in meters. The constant ε_0 is the permittivity of free space; its numerical value in SI units is $\varepsilon_0 = 8.85 \times 10^{–12}\ \text{F}/\text{m}$. The units of F/m are equivalent to $\text{C}^2/\text{N}\cdot\text{m}^2$. The small numerical value of ε_0 is related to the large size of the farad. A parallel plate capacitor must have a large area to have a capacitance approaching a farad. (Note that the above equation is valid when the parallel plates are separated by air or free space. When another material is placed between the plates, the equation is modified, as discussed below.)

Example:

Capacitance and Charge Stored in a Parallel Plate Capacitor

(a) What is the capacitance of a parallel plate capacitor with metal plates, each of area $1.00\ \text{m}^2$, separated by 1.00 mm? (b) What charge is stored in this capacitor if a voltage of 3.00×10^3 V is applied to it?

Strategy

Finding the capacitance C is a straightforward application of the equation $C=\varepsilon_0 A/d$. Once C is found, the charge stored can be found using the equation $Q = \text{CV}$.

Solution for (a)

Entering the given values into the equation for the capacitance of a parallel plate capacitor yields

Equation:

$$\begin{aligned} C &= \varepsilon_0\frac{A}{d} = \left(8.85\times 10^{–12}\,\frac{\text{F}}{\text{m}}\right)\frac{1.00\ \text{m}^2}{1.00{\times}10^{–3}\ \text{m}} \\ &= 8.85\times 10^{–9}\ \text{F} = 8.85\ \text{nF}. \end{aligned}$$

Discussion for (a)

This small value for the capacitance indicates how difficult it is to make a device with a large capacitance. Special techniques help, such as using very large area thin foils placed close together.

Solution for (b)

The charge stored in any capacitor is given by the equation $Q = \text{CV}$. Entering the known values into this equation gives

Equation:

$$\begin{array}{lll} Q & = & CV = \left(8.85 \times 10^{-9}\ \text{F}\right)\left(3.00 \times 10^{3}\ \text{V}\right) \\ & = & 26.6\ \mu\text{C}. \end{array}$$

Discussion for (b)

This charge is only slightly greater than those found in typical static electricity. Since air breaks down at about $3.00 \times 10^{6}\ \text{V/m}$, more charge cannot be stored on this capacitor by increasing the voltage.

Another interesting biological example dealing with electric potential is found in the cell's plasma membrane. The membrane sets a cell off from its surroundings and also allows ions to selectively pass in and out of the cell. There is a potential difference across the membrane of about $–70\ \text{mV}$. This is due to the mainly negatively charged ions in the cell and the predominance of positively charged sodium (Na^{+}) ions outside. Things change when a nerve cell is stimulated. Na^{+} ions are allowed to pass through the membrane into the cell, producing a positive membrane potential—the nerve signal. The cell membrane is about 7 to 10 nm thick. An approximate value of the electric field across it is given by

Equation:

$$E = \frac{V}{d} = \frac{–70 \times 10^{-3}\ \text{V}}{8 \times 10^{-9}\ \text{m}} = –9 \times 10^{6}\ \text{V/m}.$$

This electric field is enough to cause a breakdown in air.

Dielectric

The previous example highlights the difficulty of storing a large amount of charge in capacitors. If d is made smaller to produce a larger capacitance, then the maximum voltage must be reduced proportionally to avoid breakdown (since $E = V/d$). An important solution to this difficulty is to put an insulating material, called a **dielectric**, between the plates of a capacitor and allow d to be as small as possible. Not only does the smaller d make the capacitance greater, but many insulators can withstand greater electric fields than air before breaking down.

There is another benefit to using a dielectric in a capacitor. Depending on the material used, the capacitance is greater than that given by the equation $C = \varepsilon_0 \frac{A}{d}$ by a factor κ, called the *dielectric constant*. A parallel plate capacitor with a dielectric between its plates has a capacitance given by

Equation:

$$C = \kappa\varepsilon_0 \frac{A}{d} \text{ (parallel plate capacitor with dielectric).}$$

Values of the dielectric constant κ for various materials are given in [link]. Note that κ for vacuum is exactly 1, and so the above equation is valid in that case, too. If a dielectric is used, perhaps by placing Teflon between the plates of the capacitor in [link], then the capacitance is greater by the factor κ, which for Teflon is 2.1.

Note:

Take-Home Experiment: Building a Capacitor

How large a capacitor can you make using a chewing gum wrapper? The plates will be the aluminum foil, and the separation (dielectric) in between will be the paper.

Material	**Dielectric constant κ**	**Dielectric strength (V/m)**
Vacuum	1.00000	—
Air	1.00059	3×10^6
Bakelite	4.9	24×10^6
Fused quartz	3.78	8×10^6
Neoprene rubber	6.7	12×10^6
Nylon	3.4	14×10^6
Paper	3.7	16×10^6
Polystyrene	2.56	24×10^6
Pyrex glass	5.6	14×10^6
Silicon oil	2.5	15×10^6
Strontium titanate	233	8×10^6
Teflon	2.1	60×10^6
Water	80	—

Dielectric Constants and Dielectric Strengths for Various Materials at 20ºC

Note also that the dielectric constant for air is very close to 1, so that air-filled capacitors act much like those with vacuum between their plates *except* that the air can become conductive if the electric field strength becomes too great. (Recall that $E = V/d$ for a parallel plate capacitor.) Also shown in [link] are maximum electric field strengths in V/m, called **dielectric strengths**, for several materials. These are the fields above which the material begins to break down and conduct. The dielectric strength imposes a limit on the voltage that can be applied for a given plate separation. For instance, in [link], the separation is 1.00 mm, and so the voltage limit for air is

Equation:

$$\begin{array}{rcl} V & = & E \cdot d \\ & = & (3 \times 10^6\ \text{V/m})(1.00 \times 10^{-3}\ \text{m}) \\ & = & 3000\ \text{V.} \end{array}$$

However, the limit for a 1.00 mm separation filled with Teflon is 60,000 V, since the dielectric strength of Teflon is 60×10^6 V/m. So the same capacitor filled with Teflon has a greater capacitance and can be subjected to a much greater voltage. Using the capacitance we calculated in the above example for the air-filled parallel plate capacitor, we find that the Teflon-filled capacitor can store a maximum charge of

Equation:

$$\begin{array}{rcl} Q & = & CV \\ & = & \kappa C_{\text{air}} V \\ & = & (2.1)(8.85\ \text{nF})(6.0 \times 10^4\ \text{V}) \\ & = & 1.1\ \text{mC.} \end{array}$$

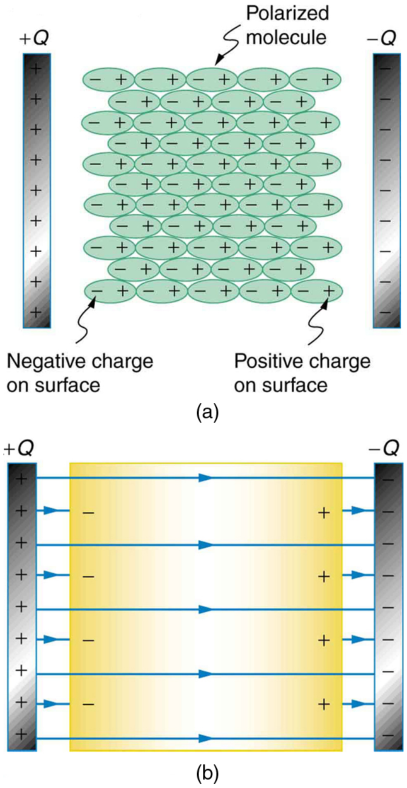
This is 42 times the charge of the same air-filled capacitor.

Note:

Dielectric Strength

The maximum electric field strength above which an insulating material begins to break down and conduct is called its dielectric strength.

Microscopically, how does a dielectric increase capacitance? Polarization of the insulator is responsible. The more easily it is polarized, the greater its dielectric constant κ. Water, for example, is a **polar molecule** because one end of the molecule has a slight positive charge and the other end has a slight negative charge. The polarity of water causes it to have a relatively large dielectric constant of 80. The effect of polarization can be best explained in terms of the characteristics of the Coulomb force. [link] shows the separation of charge schematically in the molecules of a dielectric material placed between the charged plates of a capacitor. The Coulomb force between the closest ends of the molecules and the charge on the plates is attractive and very strong, since they are very close together. This attracts more charge onto the plates than if the space were empty and the opposite charges were a distance d away.



(a) The molecules in the insulating material between the plates of a capacitor are polarized by the charged plates. This produces a layer of opposite charge on the surface of the dielectric that attracts more charge onto the plate, increasing its capacitance. (b) The dielectric reduces the electric field strength inside the capacitor, resulting in a smaller voltage between the plates for the same charge. The capacitor stores the same charge for a smaller voltage, implying that it has a larger capacitance because of the dielectric.

Another way to understand how a dielectric increases capacitance is to consider its effect on the electric field inside the capacitor. [link](b) shows the electric field lines with a dielectric in place. Since the field lines end on charges in the dielectric, there are fewer of them going from one side of the capacitor to the other. So the electric field strength is less than if there were a vacuum between the plates, even though the same charge is on the plates. The voltage between the plates is $V = \text{Ed}$, so it too is reduced by the dielectric. Thus there is a smaller voltage V for the same charge Q; since $C = Q/V$, the capacitance C is greater.

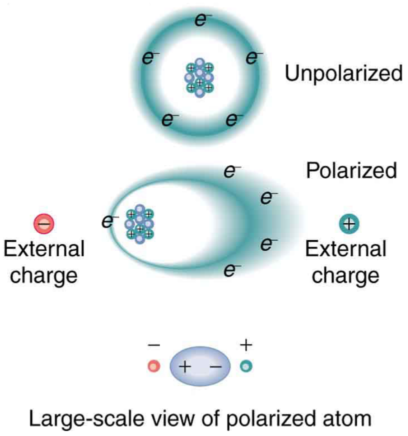
The dielectric constant is generally defined to be $\kappa = E_0/E$, or the ratio of the electric field in a vacuum to that in the dielectric material, and is intimately related to the polarizability of the material.

Note:

Things Great and Small

The Submicroscopic Origin of Polarization

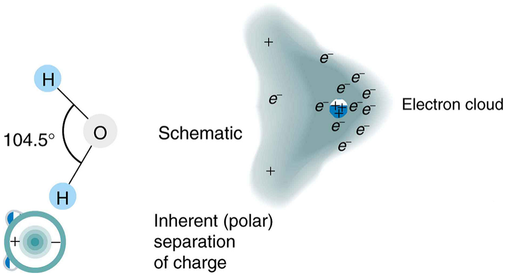
Polarization is a separation of charge within an atom or molecule. As has been noted, the planetary model of the atom pictures it as having a positive nucleus orbited by negative electrons, analogous to the planets orbiting the Sun. Although this model is not completely accurate, it is very helpful in explaining a vast range of phenomena and will be refined elsewhere, such as in Atomic Physics. The submicroscopic origin of polarization can be modeled as shown in [link].



Artist’s conception of a polarized atom. The orbits of electrons around the nucleus are shifted slightly by the external charges (shown exaggerated). The resulting separation of charge within the atom means that it is polarized. Note that the unlike charge is now closer to the external charges, causing the polarization.

We will find in Atomic Physics that the orbits of electrons are more properly viewed as electron clouds with the density of the cloud related to the probability of finding an electron in that location (as opposed to the definite locations and paths of planets in their orbits around the Sun). This cloud is shifted by the Coulomb force so that the atom on average has a separation of charge. Although the atom remains neutral, it can now be the source of a Coulomb force, since a charge brought near the atom will be closer to one type of charge than the other.

Some molecules, such as those of water, have an inherent separation of charge and are thus called polar molecules. [link] illustrates the separation of charge in a water molecule, which has two hydrogen atoms and one oxygen atom (H_2O). The water molecule is not symmetric—the hydrogen atoms are repelled to one side, giving the molecule a boomerang shape. The electrons in a water molecule are more concentrated around the more highly charged oxygen nucleus than around the hydrogen nuclei. This makes the oxygen end of the molecule slightly negative and leaves the hydrogen ends slightly positive. The inherent separation of charge in polar molecules makes it easier to align them with external fields and charges. Polar molecules therefore exhibit greater polarization effects and have greater dielectric constants. Those who study chemistry will find that the polar nature of water has many effects. For example, water molecules gather ions much more effectively because they have an electric field and a separation of charge to attract charges of both signs. Also, as brought out in the previous chapter, polar water provides a shield or screening of the electric fields in the highly charged molecules of interest in biological systems.



Artist’s conception of a water molecule. There is an inherent separation of charge, and so water is a polar molecule. Electrons in the molecule are attracted to the oxygen nucleus and leave an excess of positive charge near the two hydrogen nuclei. (Note that the schematic on the right is a rough illustration of the distribution of electrons in the water molecule. It does not show the actual numbers of protons and electrons involved in the structure.)

Note:
PhET Explorations: Capacitor Lab
Explore how a capacitor works! Change the size of the plates and add a dielectric to see the effect on capacitance. Change the voltage and see charges built up on the plates. Observe the electric field in the capacitor. Measure the voltage and the electric field.

Capacitor Lab

Section Summary

- A capacitor is a device used to store charge.
- The amount of charge Q a capacitor can store depends on two major factors—the voltage applied and the capacitor's physical characteristics, such as its size.
- The capacitance C is the amount of charge stored per volt, or
  Equation:

$$C=\frac{Q}{V}.$$

- The capacitance of a parallel plate capacitor is $C=\varepsilon_0\frac{A}{d}$, when the plates are separated by air or free space. ε_0 is called the permittivity of free space.
- A parallel plate capacitor with a dielectric between its plates has a capacitance given by
  Equation:

$$C = \kappa\varepsilon_0 \frac{A}{d},$$

where κ is the dielectric constant of the material.
- The maximum electric field strength above which an insulating material begins to break down and conduct is called dielectric strength.

Conceptual Questions

Exercise:

Problem:

Does the capacitance of a device depend on the applied voltage? What about the charge stored in it?

Exercise:

Problem:

Use the characteristics of the Coulomb force to explain why capacitance should be proportional to the plate area of a capacitor. Similarly, explain why capacitance should be inversely proportional to the separation between plates.

Exercise:

Problem:

Give the reason why a dielectric material increases capacitance compared with what it would be with air between the plates of a capacitor. What is the independent reason that a dielectric material also allows a greater voltage to be applied to a capacitor? (The dielectric thus increases C and permits a greater V.)

Exercise:

Problem:

How does the polar character of water molecules help to explain water’s relatively large dielectric constant? ([link])

Exercise:

Problem:

Sparks will occur between the plates of an air-filled capacitor at lower voltage when the air is humid than when dry. Explain why, considering the polar character of water molecules.

Exercise:

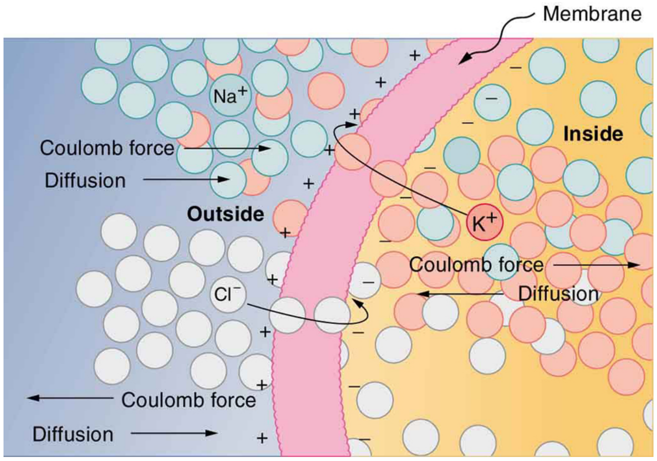
Problem:

Water has a large dielectric constant, but it is rarely used in capacitors. Explain why.

Exercise:

Problem:

Membranes in living cells, including those in humans, are characterized by a separation of charge across the membrane. Effectively, the membranes are thus charged capacitors with important functions related to the potential difference across the membrane. Is energy required to separate these charges in living membranes and, if so, is its source the metabolization of food energy or some other source?



Membrane
Na⁺
Coulomb force
Diffusion
Outside
Cl⁻
K⁺
Inside
Coulomb force
Diffusion
Coulomb force
Diffusion

The semipermeable membrane of a cell has different concentrations of ions inside and out. Diffusion moves the K^{+} (potassium) and Cl^{-} (chloride) ions in the directions shown, until the Coulomb force halts further transfer. This results in a layer of positive charge on the outside, a layer of negative charge on the inside, and thus a voltage across the cell membrane. The membrane is normally impermeable to Na^{+} (sodium ions).

Problems & Exercises

Exercise:

Problem:

What charge is stored in a $180\ \mu\mathrm{F}$ capacitor when 120 V is applied to it?

Solution:

$21.6\ \mathrm{mC}$

Exercise:

Problem:

Find the charge stored when 5.50 V is applied to an 8.00 pF capacitor.

Exercise:

Problem: What charge is stored in the capacitor in [link]?

Solution:

80.0 mC

Exercise:

Problem:

Calculate the voltage applied to a 2.00 μF capacitor when it holds 3.10 μC of charge.

Exercise:

Problem:

What voltage must be applied to an 8.00 nF capacitor to store 0.160 mC of charge?

Solution:

20.0 kV

Exercise:

Problem:

What capacitance is needed to store 3.00 μC of charge at a voltage of 120 V?

Exercise:

Problem:

What is the capacitance of a large Van de Graaff generator's terminal, given that it stores 8.00 mC of charge at a voltage of 12.0 MV?

Solution:

667 pF

Exercise:

Problem:

Find the capacitance of a parallel plate capacitor having plates of area $5.00\ \text{m}^2$ that are separated by 0.100 mm of Teflon.

Exercise:

Problem:

(a)What is the capacitance of a parallel plate capacitor having plates of area $1.50\ \text{m}^2$ that are separated by 0.0200 mm of neoprene rubber? (b) What charge does it hold when 9.00 V is applied to it?

Solution:

(a) $4.4\ \mu\text{F}$

(b) $4.0 \times 10^{-5}\ \text{C}$

Exercise:

Problem: Integrated Concepts

A prankster applies 450 V to an $80.0\ \mu\text{F}$ capacitor and then tosses it to an unsuspecting victim. The victim’s finger is burned by the discharge of the capacitor through 0.200 g of flesh. What is the temperature increase of the flesh? Is it reasonable to assume no phase change?

Exercise:

Problem:Unreasonable Results

(a) A certain parallel plate capacitor has plates of area $4.00\ \text{m}^2$, separated by 0.0100 mm of nylon, and stores 0.170 C of charge. What is the applied voltage? (b) What is unreasonable about this result? (c) Which assumptions are responsible or inconsistent?

Solution:

(a) 14.2 kV

(b) The voltage is unreasonably large, more than 100 times the breakdown voltage of nylon.

(c) The assumed charge is unreasonably large and cannot be stored in a capacitor of these dimensions.

Glossary

capacitor
a device that stores electric charge

capacitance
amount of charge stored per unit volt

dielectric
an insulating material

dielectric strength
the maximum electric field above which an insulating material begins to break down and conduct

parallel plate capacitor
two identical conducting plates separated by a distance

polar molecule
a molecule with inherent separation of charge

Capacitors in Series and Parallel

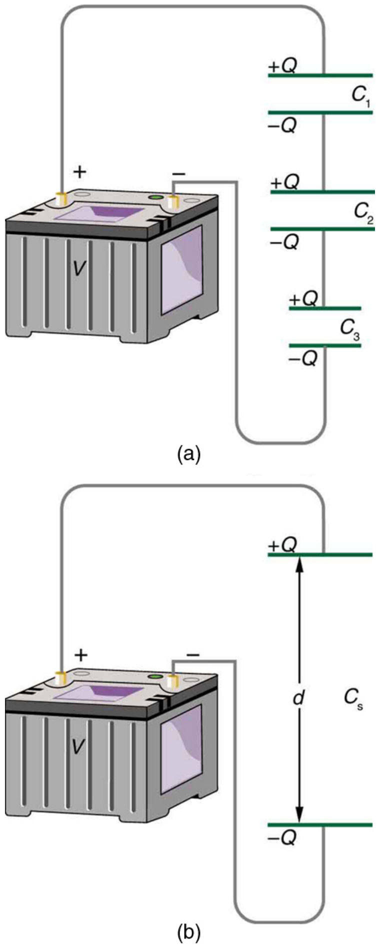
- Derive expressions for total capacitance in series and in parallel.
- Identify series and parallel parts in the combination of connection of capacitors.
- Calculate the effective capacitance in series and parallel given individual capacitances.

Several capacitors may be connected together in a variety of applications. Multiple connections of capacitors act like a single equivalent capacitor. The total capacitance of this equivalent single capacitor depends both on the individual capacitors and how they are connected. There are two simple and common types of connections, called *series* and *parallel*, for which we can easily calculate the total capacitance. Certain more complicated connections can also be related to combinations of series and parallel.

Capacitance in Series

[link](a) shows a series connection of three capacitors with a voltage applied. As for any capacitor, the capacitance of the combination is related to charge and voltage by $C=\frac{Q}{V}$.

Note in [link] that opposite charges of magnitude Q flow to either side of the originally uncharged combination of capacitors when the voltage V is applied. Conservation of charge requires that equal-magnitude charges be created on the plates of the individual capacitors, since charge is only being separated in these originally neutral devices. The end result is that the combination resembles a single capacitor with an effective plate separation greater than that of the individual capacitors alone. (See [link](b).) Larger plate separation means smaller capacitance. It is a general feature of series connections of capacitors that the total capacitance is less than any of the individual capacitances.


+Q
C1
−Q
+
−
+Q
C2
−Q
V
+Q
C3
−Q
(a)
+Q
+
−
d
Cs
V
−Q
(b)

(a) Capacitors connected in series. The magnitude of the charge on each plate is Q. (b) An equivalent capacitor has a larger plate separation d. Series connections produce a total capacitance that is less than that of

any of the
individual
capacitors.

We can find an expression for the total capacitance by considering the voltage across the individual capacitors shown in [link]. Solving $C = \frac{Q}{V}$ for V gives $V = \frac{Q}{C}$. The voltages across the individual capacitors are thus $V_1 = \frac{Q}{C_1}$, $V_2 = \frac{Q}{C_2}$, and $V_3 = \frac{Q}{C_3}$. The total voltage is the sum of the individual voltages:

Equation:

$$V = V_1 + V_2 + V_3.$$

Now, calling the total capacitance C_S for series capacitance, consider that

Equation:

$$V = \frac{Q}{C_\mathrm{S}} = V_1 + V_2 + V_3.$$

Entering the expressions for V_1, V_2, and V_3, we get

Equation:

$$\frac{Q}{C_\mathrm{S}} = \frac{Q}{C_1} + \frac{Q}{C_2} + \frac{Q}{C_3}.$$

Canceling the Qs, we obtain the equation for the total capacitance in series C_S to be

Equation:

$$\frac{1}{C_\mathrm{S}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots,$$

where “...” indicates that the expression is valid for any number of capacitors connected in series. An expression of this form always results in a total capacitance C_S that is less than any of the individual capacitances C_1, C_2, ..., as the next example illustrates.

Note:

Total Capacitance in Series, C_s

Total capacitance in series: $\frac{1}{C_\text{S}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + ...$

Example:

What Is the Series Capacitance?

Find the total capacitance for three capacitors connected in series, given their individual capacitances are 1.000, 5.000, and 8.000 µF.

Strategy

With the given information, the total capacitance can be found using the equation for capacitance in series.

Solution

Entering the given capacitances into the expression for $\frac{1}{C_\text{S}}$ gives $\frac{1}{C_S} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$.

Equation:

$$\frac{1}{C_\text{S}} = \frac{1}{1.000\ \mu\text{F}} + \frac{1}{5.000\ \mu\text{F}} + \frac{1}{8.000\ \mu\text{F}} = \frac{1.325}{\mu\text{F}}$$

Inverting to find C_S yields $C_\text{S} = \frac{\mu\text{F}}{1.325} = 0.755\ \mu\text{F}$.

Discussion

The total series capacitance C_s is less than the smallest individual capacitance, as promised. In series connections of capacitors, the sum is less than the parts. In fact, it is less than any individual. Note that it is sometimes possible, and more convenient, to solve an equation like the above by finding the least common denominator, which in this case (showing only whole-number calculations) is 40. Thus,

Equation:

$$\frac{1}{C_{\text{S}}} = \frac{40}{40\ \mu\text{F}} + \frac{8}{40\ \mu\text{F}} + \frac{5}{40\ \mu\text{F}} = \frac{53}{40\ \mu\text{F}},$$

so that

Equation:

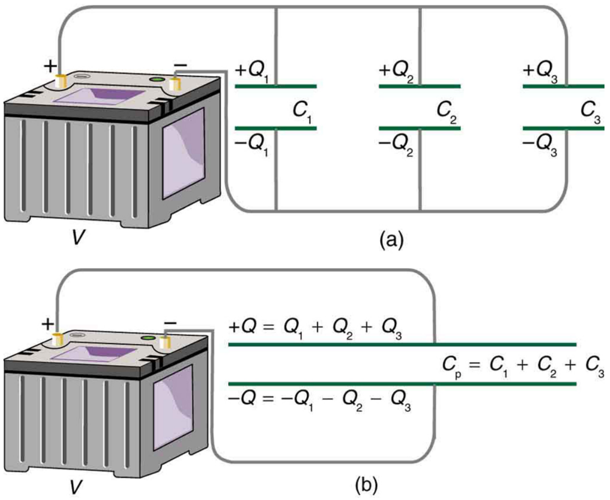
$$C_{\text{S}} = \frac{40\ \mu\text{F}}{53} = 0.755\ \mu\text{F}.$$

Capacitors in Parallel

[link](a) shows a parallel connection of three capacitors with a voltage applied. Here the total capacitance is easier to find than in the series case. To find the equivalent total capacitance C_{p}, we first note that the voltage across each capacitor is V, the same as that of the source, since they are connected directly to it through a conductor. (Conductors are equipotentials, and so the voltage across the capacitors is the same as that across the voltage source.) Thus the capacitors have the same charges on them as they would have if connected individually to the voltage source. The total charge Q is the sum of the individual charges:

Equation:

$$Q = Q_1 + Q_2 + Q_3.$$



(a) Capacitors in parallel. Each is connected directly to the voltage source just as if it were all alone, and so the total capacitance in parallel is just the sum of the individual capacitances. (b) The equivalent capacitor has a larger plate area and can therefore hold more charge than the individual capacitors.

Using the relationship $Q = \text{CV}$, we see that the total charge is $Q = C_{\text{p}}V$, and the individual charges are $Q_1 = C_1V$, $Q_2 = C_2V$, and $Q_3 = C_3V$. Entering these into the previous equation gives

Equation:

$$C_{\text{p}}V = C_1V + C_2V + C_3V.$$

Canceling V from the equation, we obtain the equation for the total capacitance in parallel C_{p}:

Equation:

$$C_{\mathrm{p}} = C_1 + C_2 + C_3 + \ldots.$$

Total capacitance in parallel is simply the sum of the individual capacitances. (Again the "..." indicates the expression is valid for any number of capacitors connected in parallel.) So, for example, if the capacitors in the example above were connected in parallel, their capacitance would be

Equation:

$$C_{\mathrm{p}} = 1.000\ \mu\mathrm{F} + 5.000\ \mu\mathrm{F} + 8.000\ \mu\mathrm{F} = 14.000\ \mu\mathrm{F}.$$

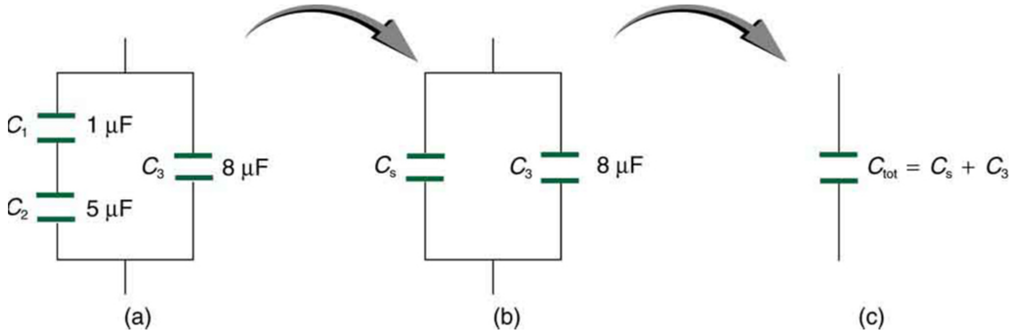
The equivalent capacitor for a parallel connection has an effectively larger plate area and, thus, a larger capacitance, as illustrated in [link](b).

Note:

Total Capacitance in Parallel, C_{p}

Total capacitance in parallel $C_{\mathrm{p}} = C_1 + C_2 + C_3 + \ldots$

More complicated connections of capacitors can sometimes be combinations of series and parallel. (See [link].) To find the total capacitance of such combinations, we identify series and parallel parts, compute their capacitances, and then find the total.



(a) This circuit contains both series and parallel connections of capacitors. See [link] for the calculation of the overall capacitance of the circuit. (b) C_1 and C_2 are in series; their equivalent capacitance C_S is less than either of them. (c) Note that C_S is in parallel with C_3. The total capacitance is, thus, the sum of C_S and C_3.

Example:

A Mixture of Series and Parallel Capacitance

Find the total capacitance of the combination of capacitors shown in [link]. Assume the capacitances in [link] are known to three decimal places ($C_1 = 1.000\ \mu\mathrm{F}$, $C_2 = 5.000\ \mu\mathrm{F}$, and $C_3 = 8.000\ \mu\mathrm{F}$), and round your answer to three decimal places.

Strategy

To find the total capacitance, we first identify which capacitors are in series and which are in parallel. Capacitors C_1 and C_2 are in series. Their combination, labeled C_S in the figure, is in parallel with C_3.

Solution

Since C_1 and C_2 are in series, their total capacitance is given by $\frac{1}{C_\mathrm{S}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$. Entering their values into the equation gives

Equation:

$$\frac{1}{C_\mathrm{S}} = \frac{1}{C_1} + \frac{1}{C_2} = \frac{1}{1.000\ \mu\mathrm{F}} + \frac{1}{5.000\ \mu\mathrm{F}} = \frac{1.200}{\mu\mathrm{F}}.$$

Inverting gives

Equation:

$$C_\mathrm{S} = 0.833\ \mu\mathrm{F}.$$

This equivalent series capacitance is in parallel with the third capacitor; thus, the total is the sum

Equation:

$$\begin{aligned} C_{\text{tot}} &= C_{\text{S}} + C_{\text{S}} \\ &= 0.833\ \mu\text{F} + 8.000\ \mu\text{F} \\ &= 8.833\ \mu\text{F}. \end{aligned}$$

Discussion
This technique of analyzing the combinations of capacitors piece by piece until a total is obtained can be applied to larger combinations of capacitors.

Section Summary

- Total capacitance in series $\frac{1}{C_{\text{S}}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots$
- Total capacitance in parallel $C_{\text{p}} = C_1 + C_2 + C_3 + \ldots$
- If a circuit contains a combination of capacitors in series and parallel, identify series and parallel parts, compute their capacitances, and then find the total.

Conceptual Questions

Exercise:

Problem:

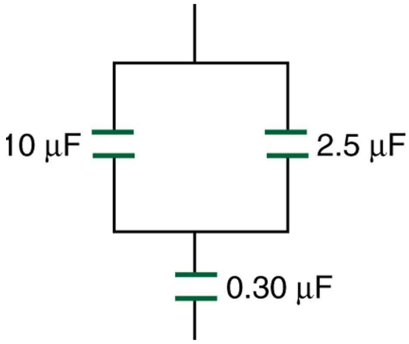
If you wish to store a large amount of energy in a capacitor bank, would you connect capacitors in series or parallel? Explain.

Problems & Exercises

Exercise:

Problem:

Find the total capacitance of the combination of capacitors in [link].



A combination of series and parallel connections of capacitors.

Solution:

0.293 μF

Exercise:

Problem:

Suppose you want a capacitor bank with a total capacitance of 0.750 F and you possess numerous 1.50 mF capacitors. What is the smallest number you could hook together to achieve your goal, and how would you connect them?

Exercise:

Problem:

What total capacitances can you make by connecting a 5.00 μF and an 8.00 μF capacitor together?

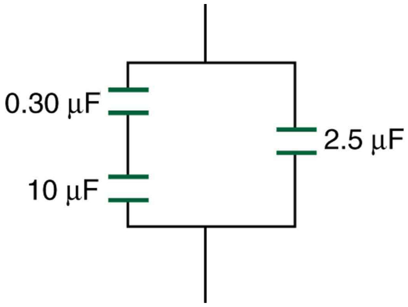
Solution:

3.08 μF in series combination, 13.0 μF in parallel combination

Exercise:

Problem:

Find the total capacitance of the combination of capacitors shown in [link].


0.30 μF
10 μF
2.5 μF

A combination of series and parallel connections of capacitors.

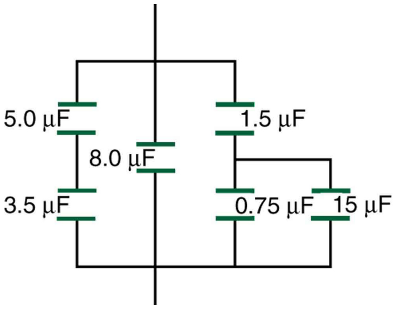
Solution:

2.79 μF

Exercise:

Problem:

Find the total capacitance of the combination of capacitors shown in [link].


5.0 μF
8.0 μF
3.5 μF
1.5 μF
0.75 μF
15 μF

A combination of series and parallel

connections of
capacitors.

Exercise:

Problem: **Unreasonable Results**

(a) An $8.00\ \mu F$ capacitor is connected in parallel to another capacitor, producing a total capacitance of $5.00\ \mu F$. What is the capacitance of the second capacitor? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $-3.00\ \mu F$

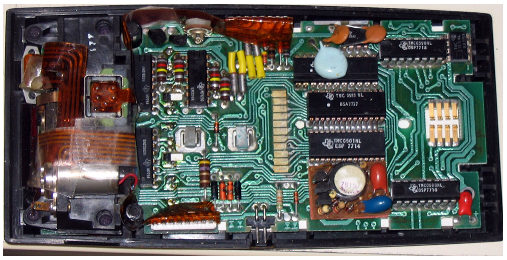
(b) You cannot have a negative value of capacitance.

(c) The assumption that the capacitors were hooked up in parallel, rather than in series, was incorrect. A parallel connection always produces a greater capacitance, while here a smaller capacitance was assumed. This could happen only if the capacitors are connected in series.

Energy Stored in Capacitors

- List some uses of capacitors.
- Express in equation form the energy stored in a capacitor.
- Explain the function of a defibrillator.

Most of us have seen dramatizations in which medical personnel use a **defibrillator** to pass an electric current through a patient's heart to get it to beat normally. (Review [link].) Often realistic in detail, the person applying the shock directs another person to "make it 400 joules this time." The energy delivered by the defibrillator is stored in a capacitor and can be adjusted to fit the situation. SI units of joules are often employed. Less dramatic is the use of capacitors in microelectronics, such as certain handheld calculators, to supply energy when batteries are charged. (See [link].) Capacitors are also used to supply energy for flash lamps on cameras.



Energy stored in the large capacitor is used to preserve the memory of an electronic calculator when its batteries are charged. (credit: Kucharek, Wikimedia Commons)

Energy stored in a capacitor is electrical potential energy, and it is thus related to the charge Q and voltage V on the capacitor. We must be careful when applying the equation for electrical potential energy $\Delta\mathrm{PE} = q\Delta V$ to a capacitor. Remember that $\Delta\mathrm{PE}$ is the potential energy of a charge q going through a voltage ΔV. But the capacitor starts with zero voltage and gradually comes up to its full voltage as it is charged. The first charge placed on a capacitor experiences a change in voltage $\Delta V = 0$, since the capacitor has zero voltage when uncharged. The final charge placed on a capacitor experiences $\Delta V = V$, since the capacitor now has its full voltage V on it. The average voltage on the capacitor during the charging process is $V/2$, and so the average voltage experienced by the full charge q is $V/2$. Thus the energy stored in a capacitor, E_{cap}, is

Equation:

$$E_{\text{cap}} = \frac{QV}{2},$$

where Q is the charge on a capacitor with a voltage V applied. (Note that the energy is not QV, but $\mathrm{QV}/2$.) Charge and voltage are related to the capacitance C of a capacitor by $Q = \mathrm{CV}$, and so the expression for E_{cap} can be algebraically manipulated into three equivalent expressions:

Equation:

$$E_{\text{cap}} = \frac{QV}{2} = \frac{CV^2}{2} = \frac{Q^2}{2C},$$

where Q is the charge and V the voltage on a capacitor C. The energy is in joules for a charge in coulombs, voltage in volts, and capacitance in farads.

Note:

Energy Stored in Capacitors

The energy stored in a capacitor can be expressed in three ways:

Equation:

$$E_{\text{cap}} = \frac{QV}{2} = \frac{CV^2}{2} = \frac{Q^2}{2C},$$

where Q is the charge, V is the voltage, and C is the capacitance of the capacitor. The energy is in joules for a charge in coulombs, voltage in volts, and capacitance in farads.

In a defibrillator, the delivery of a large charge in a short burst to a set of paddles across a person’s chest can be a lifesaver. The person’s heart attack might have arisen from the onset of fast, irregular beating of the heart—cardiac or ventricular fibrillation. The application of a large shock of electrical energy can terminate the arrhythmia and allow the body’s pacemaker to resume normal patterns. Today it is common for ambulances to carry a defibrillator, which also uses an electrocardiogram to analyze the patient’s heartbeat pattern. Automated external defibrillators (AED) are found in many public places ([link]). These are designed to be used by lay persons. The device automatically diagnoses the patient’s heart condition and then applies the shock with appropriate energy and waveform. CPR is recommended in many cases before use of an AED.

POWERHEART AED
MAY-2011

Automated external defibrillators are found in many public places. These portable units provide verbal instructions for use in the important first few minutes for a person suffering a cardiac attack. (credit: Owain Davies, Wikimedia Commons)

Example:

Capacitance in a Heart Defibrillator

A heart defibrillator delivers 4.00×10^2 J of energy by discharging a capacitor initially at 1.00×10^4 V. What is its capacitance?

Strategy

We are given E_{cap} and V, and we are asked to find the capacitance C. Of the three expressions in the equation for E_{cap}, the most convenient relationship is

Equation:

$$E_{\text{cap}} = \frac{CV^2}{2}.$$

Solution

Solving this expression for C and entering the given values yields

Equation:

$$\begin{aligned} C &= \frac{2E_{\text{cap}}}{V^2} = \frac{2(4.00 \times 10^2 \text{ J})}{(1.00 \times 10^4 \text{ V})^2} = 8.00 \times 10^{-6} \text{ F} \\ &= 8.00 \text{ μF}. \end{aligned}$$

Discussion
This is a fairly large, but manageable, capacitance at 1.00×10^4 V.

Section Summary

- Capacitors are used in a variety of devices, including defibrillators, microelectronics such as calculators, and flash lamps, to supply energy.
- The energy stored in a capacitor can be expressed in three ways:
  Equation:

  $$E_{\text{cap}} = \frac{\text{QV}}{2} = \frac{CV^2}{2} = \frac{Q^2}{2C},$$

  where Q is the charge, V is the voltage, and C is the capacitance of the capacitor. The energy is in joules when the charge is in coulombs, voltage is in volts, and capacitance is in farads.

Conceptual Questions

Exercise:

Problem:

How does the energy contained in a charged capacitor change when a dielectric is inserted, assuming the capacitor is isolated and its charge is constant? Does this imply that work was done?

Exercise:

Problem:

What happens to the energy stored in a capacitor connected to a battery when a dielectric is inserted? Was work done in the process?

Problems & Exercises

Exercise:

Problem:

(a) What is the energy stored in the $10.0\ \mu\text{F}$ capacitor of a heart defibrillator charged to $9.00 \times 10^3\ \text{V}$? (b) Find the amount of stored charge.

Solution:

(a) $405\ \text{J}$

(b) $90.0\ \text{mC}$

Exercise:

Problem:

In open heart surgery, a much smaller amount of energy will defibrillate the heart. (a) What voltage is applied to the $8.00\ \mu\text{F}$ capacitor of a heart defibrillator that stores 40.0 J of energy? (b) Find the amount of stored charge.

Solution:

(a) 3.16 kV

(b) 25.3 mC

Exercise:

Problem:

A $165\ \mu\text{F}$ capacitor is used in conjunction with a motor. How much energy is stored in it when 119 V is applied?

Exercise:

Problem:

Suppose you have a 9.00 V battery, a $2.00\ \mu\text{F}$ capacitor, and a $7.40\ \mu\text{F}$ capacitor. (a) Find the charge and energy stored if the capacitors are connected to the battery in series. (b) Do the same for a parallel connection.

Solution:

(a) $1.42 \times 10^{-5}\ \text{C}, 6.38 \times 10^{-5}\ \text{J}$

(b) $8.46 \times 10^{-5}\ \text{C}, 3.81 \times 10^{-4}\ \text{J}$

Exercise:

Problem:

A nervous physicist worries that the two metal shelves of his wood frame bookcase might obtain a high voltage if charged by static electricity, perhaps produced by friction. (a) What is the capacitance of the empty shelves if they have area $1.00 \times 10^{2}\ \text{m}^{2}$ and are 0.200 m apart? (b) What is the voltage between them if opposite charges of magnitude 2.00 nC are placed on them? (c) To show that this voltage poses a small hazard, calculate the energy stored.

Solution:

(a) $4.43 \times 10^{-12}\ \text{F}$

(b) $452\ \text{V}$

(c) $4.52 \times 10^{-7}\ \text{J}$

Exercise:

Problem:

Show that for a given dielectric material the maximum energy a parallel plate capacitor can store is directly proportional to the volume of dielectric ($\text{Volume} = A \cdot d$). Note that the applied voltage is limited by the dielectric strength.

Exercise:

Problem: Construct Your Own Problem

Consider a heart defibrillator similar to that discussed in [link]. Construct a problem in which you examine the charge stored in the capacitor of a defibrillator as a function of stored energy. Among the things to be considered are the applied voltage and whether it should vary with energy to be delivered, the range of energies involved, and the capacitance of the defibrillator. You may also wish to consider the much smaller energy needed for defibrillation during open-heart surgery as a variation on this problem.

Exercise:

Problem: Unreasonable Results

(a) On a particular day, it takes $9.60 \times 10^3\ \text{J}$ of electric energy to start a truck's engine. Calculate the capacitance of a capacitor that could store that amount of energy at 12.0 V. (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Solution:

(a) $133\ \text{F}$

(b) Such a capacitor would be too large to carry with a truck. The size of the capacitor would be enormous.

(c) It is unreasonable to assume that a capacitor can store the amount of energy needed.

Glossary

defibrillator

a machine used to provide an electrical shock to a heart attack victim's heart in order to restore the heart's normal rhythmic pattern

Introduction to Electric Current, Resistance, and Ohm's Law class="introduction"

Electric energy in massive quantities is transmitted from this hydroelectric facility, the Srisailam power station located along the Krishna River in India, by the movement of charge—that is, by electric current. (credit: Chintohere, Wikimedia Commons)

The flicker of numbers on a handheld calculator, nerve impulses carrying signals of vision to the brain, an ultrasound device sending a signal to a computer screen, the brain sending a message for a baby to twitch its toes, an electric train pulling its load over a mountain pass, a hydroelectric plant sending energy to metropolitan and rural users—these and many other examples of electricity involve *electric current*, *the movement of charge*. Humankind has indeed harnessed electricity, the basis of technology, to improve our quality of life. Whereas the previous two chapters concentrated on static electricity and the fundamental force underlying its behavior, the next few chapters will be devoted to electric and magnetic phenomena involving current. In addition to exploring applications of electricity, we shall gain new insights into nature—in particular, the fact that all magnetism results from electric current.

Current

- Define electric current, ampere, and drift velocity
- Describe the direction of charge flow in conventional current.
- Use drift velocity to calculate current and vice versa.

Electric Current

Electric current is defined to be the rate at which charge flows. A large current, such as that used to start a truck engine, moves a large amount of charge in a small time, whereas a small current, such as that used to operate a hand-held calculator, moves a small amount of charge over a long period of time. In equation form, **electric current** I is defined to be

Equation:

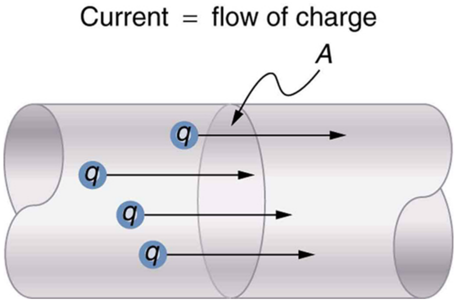
$$I=\frac{\Delta Q}{\Delta t},$$

where ΔQ is the amount of charge passing through a given area in time Δt. (As in previous chapters, initial time is often taken to be zero, in which case $\Delta t=t$.) (See [link].) The SI unit for current is the **ampere** (A), named for the French physicist André-Marie Ampère (1775–1836). Since $I=\Delta Q/\Delta t$, we see that an ampere is one coulomb per second:

Equation:

$$1\text{ A}=1\text{ C/s}$$

Not only are fuses and circuit breakers rated in amperes (or amps), so are many electrical appliances.



The rate of flow of charge is current. An ampere is the flow of one coulomb through an area in one second.

Example:

Calculating Currents: Current in a Truck Battery and a Handheld Calculator

(a) What is the current involved when a truck battery sets in motion 720 C of charge in 4.00 s while starting an engine? (b) How long does it take 1.00 C of charge to flow through a handheld calculator if a 0.300-mA current is flowing?

Strategy

We can use the definition of current in the equation $I = \Delta Q/\Delta t$ to find the current in part (a), since charge and time are given. In part (b), we rearrange the definition of current and use the given values of charge and current to find the time required.

Solution for (a)

Entering the given values for charge and time into the definition of current gives

Equation:

$$\begin{aligned} I &= \frac{\Delta Q}{\Delta t} = \frac{720\text{ C}}{4.00\text{ s}} = 180\text{ C/s} \\ &= 180\text{ A.} \end{aligned}$$

Discussion for (a)
This large value for current illustrates the fact that a large charge is moved in a small amount of time. The currents in these “starter motors” are fairly large because large frictional forces need to be overcome when setting something in motion.

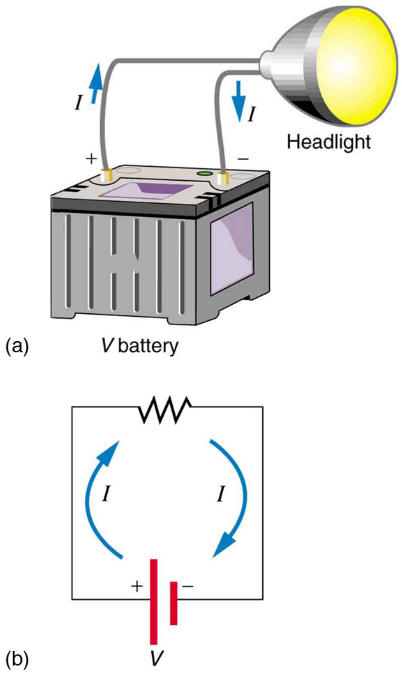
Solution for (b)
Solving the relationship $I = \Delta Q / \Delta t$ for time Δt, and entering the known values for charge and current gives

Equation:

$$\begin{aligned}\Delta t &= \frac{\Delta Q}{I} = \frac{1.00\text{ C}}{0.300 \times 10^{-3}\text{ C/s}} \\ &= 3.33 \times 10^{3}\text{ s.}\end{aligned}$$

Discussion for (b)
This time is slightly less than an hour. The small current used by the hand-held calculator takes a much longer time to move a smaller charge than the large current of the truck starter. So why can we operate our calculators only seconds after turning them on? It’s because calculators require very little energy. Such small current and energy demands allow handheld calculators to operate from solar cells or to get many hours of use out of small batteries. Remember, calculators do not have moving parts in the same way that a truck engine has with cylinders and pistons, so the technology requires smaller currents.

[link] shows a simple circuit and the standard schematic representation of a battery, conducting path, and load (a resistor). Schematics are very useful in visualizing the main features of a circuit. A single schematic can represent a wide variety of situations. The schematic in [link] (b), for example, can represent anything from a truck battery connected to a headlight lighting the street in front of the truck to a small battery connected to a penlight lighting a keyhole in a door. Such schematics are useful because the analysis is the same for a wide variety of situations. We need to understand a few schematics to apply the concepts and analysis to many more situations.



(a) A simple electric circuit. A closed path for current to flow through is supplied by conducting wires connecting a load to the terminals of a battery. (b) In this schematic, the battery is represented by the two parallel red lines, conducting wires are shown as straight lines, and the zigzag represents the load. The schematic represents a wide

variety of similar
circuits.

Note that the direction of current flow in [link] is from positive to negative. *The direction of conventional current is the direction that positive charge would flow*. Depending on the situation, positive charges, negative charges, or both may move. In metal wires, for example, current is carried by electrons—that is, negative charges move. In ionic solutions, such as salt water, both positive and negative charges move. This is also true in nerve cells. A Van de Graaff generator used for nuclear research can produce a current of pure positive charges, such as protons. [link] illustrates the movement of charged particles that compose a current. The fact that conventional current is taken to be in the direction that positive charge would flow can be traced back to American politician and scientist Benjamin Franklin in the 1700s. He named the type of charge associated with electrons negative, long before they were known to carry current in so many situations. Franklin, in fact, was totally unaware of the small-scale structure of electricity.

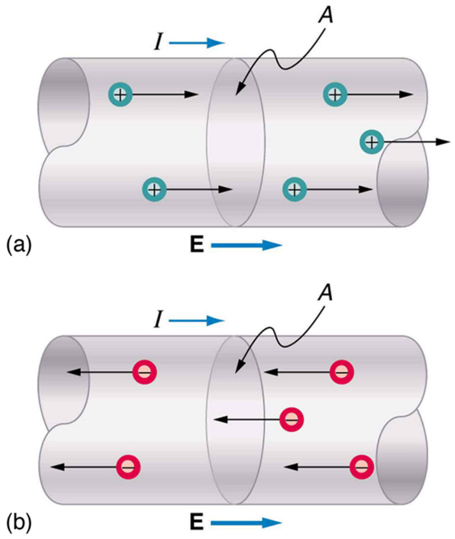
It is important to realize that there is an electric field in conductors responsible for producing the current, as illustrated in [link]. Unlike static electricity, where a conductor in equilibrium cannot have an electric field in it, conductors carrying a current have an electric field and are not in static equilibrium. An electric field is needed to supply energy to move the charges.

Note:

Making Connections: Take-Home Investigation—Electric Current Illustration

Find a straw and little peas that can move freely in the straw. Place the straw flat on a table and fill the straw with peas. When you pop one pea in at one end, a different pea should pop out the other end. This demonstration is an analogy for an electric current. Identify what compares

to the electrons and what compares to the supply of energy. What other analogies can you find for an electric current?
Note that the flow of peas is based on the peas physically bumping into each other; electrons flow due to mutually repulsive electrostatic forces.



Current I is the rate at which charge moves through an area A, such as the cross-section of a wire. Conventional current is defined to move in the direction of the electric field. (a) Positive charges move in the direction of the electric field and the same direction as conventional current. (b) Negative charges move in the direction opposite to the electric field. Conventional current is in the direction opposite to the movement of negative charge. The flow of electrons is sometimes referred to as electronic flow.

Example:
Calculating the Number of Electrons that Move through a Calculator

If the 0.300-mA current through the calculator mentioned in the [link] example is carried by electrons, how many electrons per second pass through it?

Strategy

The current calculated in the previous example was defined for the flow of positive charge. For electrons, the magnitude is the same, but the sign is opposite, $I_{\text{electrons}} = -0.300 \times 10^{-3}\ \text{C/s}$ .Since each electron $\left(e^{-}\right)$ has a charge of $–1.60 \times 10^{-19}$ C, we can convert the current in coulombs per second to electrons per second.

Solution

Starting with the definition of current, we have

Equation:

$$I_{\text{electrons}} = \frac{\Delta Q_{\text{electrons}}}{\Delta t} = \frac{-0.300 \times 10^{-3}\ \text{C}}{\text{s}}.$$

We divide this by the charge per electron, so that

Equation:

$$\begin{array}{lll}\frac{e^{-}}{\text{s}} & = & \frac{–0.300\times 10^{-3}\ \text{C}}{\text{s}}\times \frac{1\ e^{-}}{–1.60\times 10^{-19}\ \text{C}}\\ & = & 1.88\times 10^{15}\ \frac{e^{-}}{\text{s}}.\end{array}$$

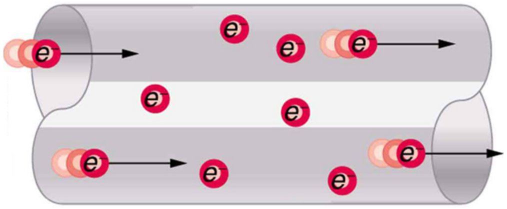
Discussion

There are so many charged particles moving, even in small currents, that individual charges are not noticed, just as individual water molecules are not noticed in water flow. Even more amazing is that they do not always keep moving forward like soldiers in a parade. Rather they are like a crowd of people with movement in different directions but a general trend to move forward. There are lots of collisions with atoms in the metal wire and, of course, with other electrons.

Drift Velocity

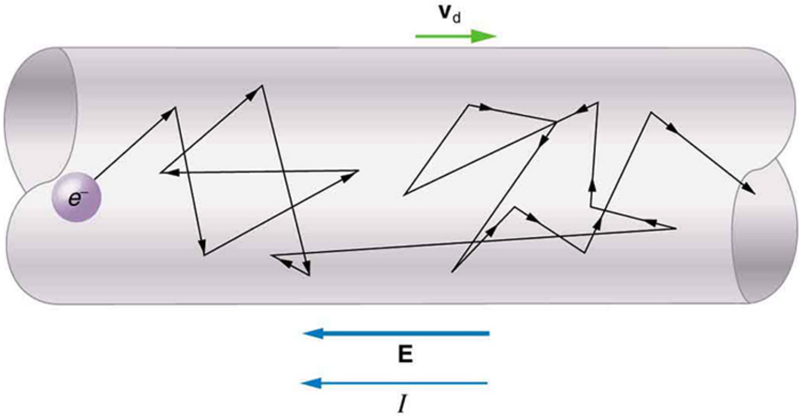
Electrical signals are known to move very rapidly. Telephone conversations carried by currents in wires cover large distances without noticeable delays. Lights come on as soon as a switch is flicked. Most electrical signals carried by currents travel at speeds on the order of $10^8\ \text{m/s}$, a significant fraction of the speed of light. Interestingly, the individual charges that make up the current move *much* more slowly on average, typically drifting at speeds on the order of $10^{-4}\ \text{m/s}$. How do we reconcile these two speeds, and what does it tell us about standard conductors?

The high speed of electrical signals results from the fact that the force between charges acts rapidly at a distance. Thus, when a free charge is forced into a wire, as in [link], the incoming charge pushes other charges ahead of it, which in turn push on charges farther down the line. The density of charge in a system cannot easily be increased, and so the signal is passed on rapidly. The resulting electrical shock wave moves through the system at nearly the speed of light. To be precise, this rapidly moving signal or shock wave is a rapidly propagating change in electric field.



When charged particles are forced into this volume of a conductor, an equal number are quickly forced to leave. The repulsion between like charges makes it difficult to increase the number of charges in a volume. Thus, as one charge enters, another leaves almost immediately, carrying the signal rapidly forward.

Good conductors have large numbers of free charges in them. In metals, the free charges are free electrons. [link] shows how free electrons move through an ordinary conductor. The distance that an individual electron can move between collisions with atoms or other electrons is quite small. The electron paths thus appear nearly random, like the motion of atoms in a gas. But there is an electric field in the conductor that causes the electrons to drift in the direction shown (opposite to the field, since they are negative). The **drift velocity** v_{d} is the average velocity of the free charges. Drift velocity is quite small, since there are so many free charges. If we have an estimate of the density of free electrons in a conductor, we can calculate the drift velocity for a given current. The larger the density, the lower the velocity required for a given current.



Free electrons moving in a conductor make many collisions with other electrons and atoms. The path of one electron is shown. The average velocity of the free charges is called the drift velocity, v_d, and it is in the direction opposite to the electric field for electrons. The collisions normally transfer energy to the conductor, requiring a constant supply of energy to maintain a steady current.

Note:
Conduction of Electricity and Heat
Good electrical conductors are often good heat conductors, too. This is because large numbers of free electrons can carry electrical current and can transport thermal energy.

The free-electron collisions transfer energy to the atoms of the conductor. The electric field does work in moving the electrons through a distance, but that work does not increase the kinetic energy (nor speed, therefore) of the electrons. The work is transferred to the conductor's atoms, possibly increasing temperature. Thus a continuous power input is required to keep a current flowing. An exception, of course, is found in superconductors, for reasons we shall explore in a later chapter. Superconductors can have a steady current without a continual supply of energy—a great energy savings. In contrast, the supply of energy can be useful, such as in a lightbulb filament. The supply of energy is necessary to increase the temperature of the tungsten filament, so that the filament glows.

Note:
Making Connections: Take-Home Investigation—Filament Observations
Find a lightbulb with a filament. Look carefully at the filament and describe its structure. To what points is the filament connected?

We can obtain an expression for the relationship between current and drift velocity by considering the number of free charges in a segment of wire, as illustrated in [link]. *The number of free charges per unit volume* is given the symbol n and depends on the material. The shaded segment has a volume Ax, so that the number of free charges in it is nAx. The charge ΔQ in this segment is thus qnAx, where q is the amount of charge on each carrier. (Recall that for electrons, q is $-1.60\times 10^{-19}\text{ C}$.) Current is charge moved per unit time; thus, if all the original charges move out of this segment in time Δt, the current is

Equation:

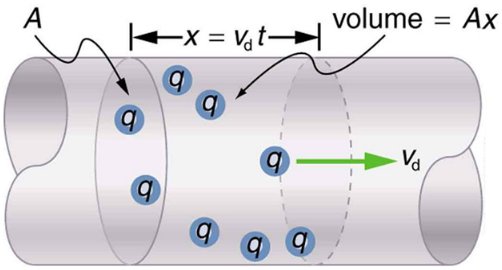
$$I=\frac{\Delta Q}{\Delta t}=\frac{\text{qnAx}}{\Delta t}.$$

Note that $x/\Delta t$ is the magnitude of the drift velocity, v_{d}, since the charges move an average distance x in a time Δt. Rearranging terms gives

Equation:

$$I=\text{nqAv}_{\text{d}},$$

where I is the current through a wire of cross-sectional area A made of a material with a free charge density n. The carriers of the current each have charge q and move with a drift velocity of magnitude v_{d}.



All the charges in the shaded volume of this wire move out in a time t, having a drift velocity of magnitude $v_{\mathrm{d}} = x/t$. See text for further discussion.

Note that simple drift velocity is not the entire story. The speed of an electron is much greater than its drift velocity. In addition, not all of the electrons in a conductor can move freely, and those that do might move somewhat faster or slower than the drift velocity. So what do we mean by free electrons? Atoms in a metallic conductor are packed in the form of a lattice structure. Some electrons are far enough away from the atomic nuclei that they do not experience the attraction of the nuclei as much as the inner electrons do. These are the free electrons. They are not bound to a single atom but can instead move freely among the atoms in a “sea” of electrons. These free electrons respond by accelerating when an electric field is applied. Of course as they move they collide with the atoms in the lattice and other electrons, generating thermal energy, and the conductor gets warmer. In an insulator, the organization of the atoms and the structure do not allow for such free electrons.

Example:

Calculating Drift Velocity in a Common Wire

Calculate the drift velocity of electrons in a 12-gauge copper wire (which has a diameter of 2.053 mm) carrying a 20.0-A current, given that there is one free electron per copper atom. (Household wiring often contains 12-gauge copper wire, and the maximum current allowed in such wire is usually 20 A.) The density of copper is $8.80 \times 10^3 \text{ kg/m}^3$.

Strategy

We can calculate the drift velocity using the equation $I = nqAv_\text{d}$. The current $I = 20.0 \text{ A}$ is given, and $q = -1.60 \times 10^{-19}\text{C}$ is the charge of an electron. We can calculate the area of a cross-section of the wire using the formula $A = \pi r^2$, where r is one-half the given diameter, 2.053 mm. We are given the density of copper, $8.80 \times 10^3 \text{ kg/m}^3$, and the periodic table shows that the atomic mass of copper is 63.54 g/mol. We can use these two quantities along with Avogadro's number, $6.02 \times 10^{23} \text{ atoms/mol}$, to determine n, the number of free electrons per cubic meter.

Solution

First, calculate the density of free electrons in copper. There is one free electron per copper atom. Therefore, is the same as the number of copper atoms per m^3. We can now find n as follows:

Equation:

$$\begin{aligned} n &= \frac{1\ e^-}{\text{atom}} \times \frac{6.02\times 10^{23} \text{ atoms}}{\text{mol}} \times \frac{1 \text{ mol}}{63.54 \text{ g}} \times \frac{1000 \text{ g}}{\text{kg}} \times \frac{8.80\times 10^3 \text{ kg}}{1 \text{ m}^3} \\ &= 8.342 \times 10^{28}\ e^-/\text{m}^3. \end{aligned}$$

The cross-sectional area of the wire is

Equation:

$$\begin{aligned} A &= \pi r^2 \\ &= \pi \left(\frac{2.053\times 10^{-3} \text{ m}}{2}\right)^2 \\ &= 3.310 \times 10^{-6} \text{ m}^2. \end{aligned}$$

Rearranging $I = nqAv_\text{d}$ to isolate drift velocity gives

Equation:

$$
\begin{aligned}
v_{\mathrm{d}} &= \frac{I}{nqA} \\
&= \frac{20.0\ \mathrm{A}}{(8.342\times10^{28}/\mathrm{m}^3)(-1.60\times10^{-19}\ \mathrm{C})(3.310\times10^{-6}\ \mathrm{m}^2)} \\
&= -4.53\times10^{-4}\ \mathrm{m/s}.
\end{aligned}
$$

Discussion

The minus sign indicates that the negative charges are moving in the direction opposite to conventional current. The small value for drift velocity (on the order of $10^{-4}\ \mathrm{m/s}$) confirms that the signal moves on the order of 10^{12} times faster (about $10^{8}\ \mathrm{m/s}$) than the charges that carry it.

Section Summary

- Electric current I is the rate at which charge flows, given by **Equation:**

  $$I = \frac{\Delta Q}{\Delta t},$$

  where ΔQ is the amount of charge passing through an area in time Δt.
- The direction of conventional current is taken as the direction in which positive charge moves.
- The SI unit for current is the ampere (A), where $1\ \mathrm{A} = 1\ \mathrm{C/s}$.
- Current is the flow of free charges, such as electrons and ions.
- Drift velocity v_{d} is the average speed at which these charges move.
- Current I is proportional to drift velocity v_{d}, as expressed in the relationship $I = \mathrm{nqAv}_{\mathrm{d}}$. Here, I is the current through a wire of cross-sectional area A. The wire's material has a free-charge density n, and each carrier has charge q and a drift velocity v_{d}.
- Electrical signals travel at speeds about 10^{12} times greater than the drift velocity of free electrons.

Conceptual Questions

Exercise:

Problem:

Can a wire carry a current and still be neutral—that is, have a total charge of zero? Explain.

Exercise:

Problem:

Car batteries are rated in ampere-hours $(\mathrm{A} \cdot \mathrm{h})$. To what physical quantity do ampere-hours correspond (voltage, charge, . . .), and what relationship do ampere-hours have to energy content?

Exercise:

Problem:

If two different wires having identical cross-sectional areas carry the same current, will the drift velocity be higher or lower in the better conductor? Explain in terms of the equation $v_{\mathrm{d}} = \frac{I}{\mathrm{nqA}}$, by considering how the density of charge carriers n relates to whether or not a material is a good conductor.

Exercise:

Problem:

Why are two conducting paths from a voltage source to an electrical device needed to operate the device?

Exercise:

Problem:

In cars, one battery terminal is connected to the metal body. How does this allow a single wire to supply current to electrical devices rather than two wires?

Exercise:

Problem:

Why isn’t a bird sitting on a high-voltage power line electrocuted? Contrast this with the situation in which a large bird hits two wires simultaneously with its wings.

Problems & Exercises

Exercise:

Problem:

What is the current in milliamperes produced by the solar cells of a pocket calculator through which 4.00 C of charge passes in 4.00 h?

Solution:

0.278 mA

Exercise:

Problem:

A total of 600 C of charge passes through a flashlight in 0.500 h. What is the average current?

Exercise:

Problem:

What is the current when a typical static charge of $0.250\ \mu\text{C}$ moves from your finger to a metal doorknob in $1.00\ \mu\text{s}$?

Solution:

0.250 A

Exercise:

Problem:

Find the current when 2.00 nC jumps between your comb and hair over a 0.500 - μs time interval.

Exercise:

Problem:

A large lightning bolt had a 20,000-A current and moved 30.0 C of charge. What was its duration?

Solution:

1.50ms

Exercise:

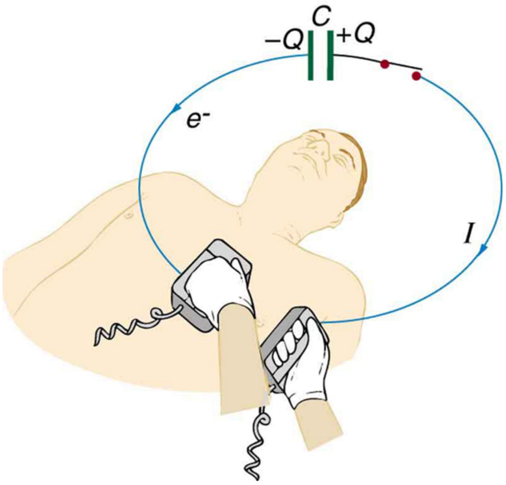
Problem:

The 200-A current through a spark plug moves 0.300 mC of charge. How long does the spark last?

Exercise:

Problem:

(a) A defibrillator sends a 6.00-A current through the chest of a patient by applying a 10,000-V potential as in the figure below. What is the resistance of the path? (b) The defibrillator paddles make contact with the patient through a conducting gel that greatly reduces the path resistance. Discuss the difficulties that would ensue if a larger voltage were used to produce the same current through the patient, but with the path having perhaps 50 times the resistance. (Hint: The current must be about the same, so a higher voltage would imply greater power. Use this equation for power: $P = I^2 R$.)



The capacitor in a defibrillation unit drives a current through the heart of a patient.

Solution:

(a) $1.67\text{k}\Omega$

(b) If a 50 times larger resistance existed, keeping the current about the same, the power would be increased by a factor of about 50 (based on the equation $P=I^2R$), causing much more energy to be transferred to the skin, which could cause serious burns. The gel used reduces the resistance, and therefore reduces the power transferred to the skin.

Exercise:

Problem:

During open-heart surgery, a defibrillator can be used to bring a patient out of cardiac arrest. The resistance of the path is $500\ \Omega$ and a 10.0-mA current is needed. What voltage should be applied?

Exercise:

Problem:

(a) A defibrillator passes 12.0 A of current through the torso of a person for 0.0100 s. How much charge moves? (b) How many electrons pass through the wires connected to the patient? (See figure two problems earlier.)

Solution:

(a) 0.120 C

(b) $7.50 \times 10^{17} \text{electrons}$

Exercise:

Problem:

A clock battery wears out after moving 10,000 C of charge through the clock at a rate of 0.500 mA. (a) How long did the clock run? (b) How many electrons per second flowed?

Exercise:

Problem:

The batteries of a submerged non-nuclear submarine supply 1000 A at full speed ahead. How long does it take to move Avogadro's number (6.02×10^{23}) of electrons at this rate?

Solution:

96.3 s

Exercise:

Problem:

Electron guns are used in X-ray tubes. The electrons are accelerated through a relatively large voltage and directed onto a metal target, producing X-rays. (a) How many electrons per second strike the target if the current is 0.500 mA? (b) What charge strikes the target in 0.750 s?

Exercise:

Problem:

A large cyclotron directs a beam of He^{++} nuclei onto a target with a beam current of 0.250 mA. (a) How many He^{++} nuclei per second is this? (b) How long does it take for 1.00 C to strike the target? (c) How long before 1.00 mol of He^{++} nuclei strike the target?

Solution:

(a) $7.81 \times 10^{14}\ \text{He}^{++}\ \text{nuclei/s}$

(b) 4.00×10^{3} s

(c) 7.71×10^{8} s

Exercise:

Problem:

Repeat the above example on [link], but for a wire made of silver and given there is one free electron per silver atom.

Exercise:

Problem:

Using the results of the above example on [link], find the drift velocity in a copper wire of twice the diameter and carrying 20.0 A.

Solution:

$-1.13\times10^{-4}\text{m/s}$

Exercise:

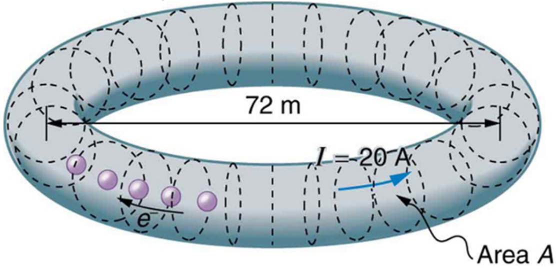
Problem:

A 14-gauge copper wire has a diameter of 1.628 mm. What magnitude current flows when the drift velocity is 1.00 mm/s? (See above example on [link] for useful information.)

Exercise:

Problem:

SPEAR, a storage ring about 72.0 m in diameter at the Stanford Linear Accelerator (closed in 2009), has a 20.0-A circulating beam of electrons that are moving at nearly the speed of light. (See [link].) How many electrons are in the beam?



Electrons circulating in the storage ring called SPEAR constitute a 20.0-A current. Because they travel close to the speed of light, each electron completes many orbits in each second.

Solution:

$9.42\times10^{13}\text{electrons}$

Glossary

electric current

the rate at which charge flows, $I = \Delta Q/\Delta t$

ampere

(amp) the SI unit for current; 1 A = 1 C/s

drift velocity

the average velocity at which free charges flow in response to an electric field

Ohm’s Law: Resistance and Simple Circuits

- Explain the origin of Ohm’s law.
- Calculate voltages, currents, or resistances with Ohm’s law.
- Explain what an ohmic material is.
- Describe a simple circuit.

What drives current? We can think of various devices—such as batteries, generators, wall outlets, and so on—which are necessary to maintain a current. All such devices create a potential difference and are loosely referred to as voltage sources. When a voltage source is connected to a conductor, it applies a potential difference V that creates an electric field. The electric field in turn exerts force on charges, causing current.

Ohm’s Law

The current that flows through most substances is directly proportional to the voltage V applied to it. The German physicist Georg Simon Ohm (1787–1854) was the first to demonstrate experimentally that the current in a metal wire is *directly proportional to the voltage applied*:

Equation:

$$I \propto V.$$

This important relationship is known as **Ohm’s law**. It can be viewed as a cause-and-effect relationship, with voltage the cause and current the effect. This is an empirical law like that for friction—an experimentally observed phenomenon. Such a linear relationship doesn’t always occur.

Resistance and Simple Circuits

If voltage drives current, what impedes it? The electric property that impedes current (crudely similar to friction and air resistance) is called **resistance** R. Collisions of moving charges with atoms and molecules in a substance transfer energy to the substance and limit current. Resistance is defined as inversely proportional to current, or

Equation:

$$I \propto \frac{1}{R}.$$

Thus, for example, current is cut in half if resistance doubles. Combining the relationships of current to voltage and current to resistance gives

Equation:

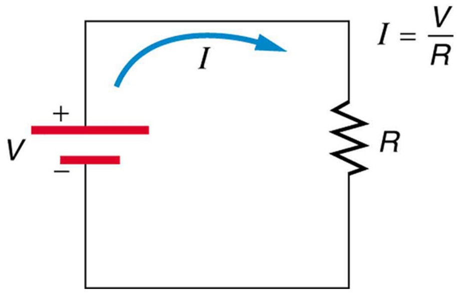
$$I = \frac{V}{R}.$$

This relationship is also called Ohm’s law. Ohm’s law in this form really defines resistance for certain materials. Ohm’s law (like Hooke’s law) is not universally valid. The many substances for which Ohm’s law holds are called **ohmic**. These include good conductors like copper and aluminum, and some poor conductors under certain circumstances. Ohmic materials have a resistance R that is independent of voltage V and current I. An object that has simple resistance is called a *resistor*, even if its resistance is small. The unit for resistance is an **ohm** and is given the symbol Ω (upper case Greek omega). Rearranging $I = \text{V/R}$ gives $R = \text{V/I}$, and so the units of resistance are 1 ohm = 1 volt per ampere:

Equation:

$$1\;\Omega = 1\frac{V}{A}.$$

[link] shows the schematic for a simple circuit. A **simple circuit** has a single voltage source and a single resistor. The wires connecting the voltage source to the resistor can be assumed to have negligible resistance, or their resistance can be included in R.



A simple electric circuit in which a closed path for current to flow is supplied by conductors (usually metal wires) connecting a load to the terminals of a battery, represented by the red parallel lines. The zigzag symbol represents the single resistor and includes any resistance in the connections to the voltage source.

Example:
Calculating Resistance: An Automobile Headlight

What is the resistance of an automobile headlight through which 2.50 A flows when 12.0 V is applied to it?

Strategy

We can rearrange Ohm’s law as stated by $I = \mathrm{V}/\mathrm{R}$ and use it to find the resistance.

Solution

Rearranging $I = \mathrm{V}/\mathrm{R}$ and substituting known values gives

Equation:

$$R=\frac{V}{I}=\frac{12.0\text{ V}}{2.50\text{ A}}=4.80\ \Omega.$$

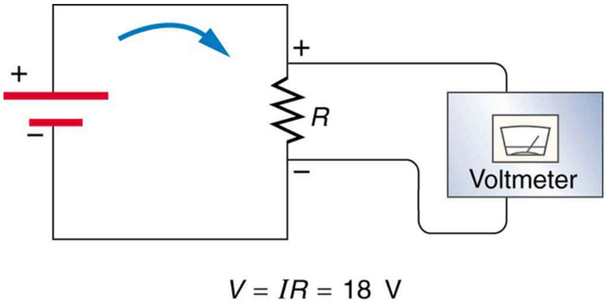
Discussion
This is a relatively small resistance, but it is larger than the cold resistance of the headlight. As we shall see in Resistance and Resistivity, resistance usually increases with temperature, and so the bulb has a lower resistance when it is first switched on and will draw considerably more current during its brief warm-up period.

Resistances range over many orders of magnitude. Some ceramic insulators, such as those used to support power lines, have resistances of $10^{12}\ \Omega$ or more. A dry person may have a hand-to-foot resistance of $10^{5}\ \Omega$, whereas the resistance of the human heart is about $10^{3}\ \Omega$. A meter-long piece of large-diameter copper wire may have a resistance of $10^{-5}\ \Omega$, and superconductors have no resistance at all (they are non-ohmic). Resistance is related to the shape of an object and the material of which it is composed, as will be seen in Resistance and Resistivity.

Additional insight is gained by solving $I=\text{V/R}$ for V, yielding
Equation:

$$V=\text{IR}.$$

This expression for V can be interpreted as the *voltage drop across a resistor produced by the flow of current* I. The phrase IR *drop* is often used for this voltage. For instance, the headlight in [link] has an IR drop of 12.0 V. If voltage is measured at various points in a circuit, it will be seen to increase at the voltage source and decrease at the resistor. Voltage is similar to fluid pressure. The voltage source is like a pump, creating a pressure difference, causing current—the flow of charge. The resistor is like a pipe that reduces pressure and limits flow because of its resistance. Conservation of energy has important consequences here. The voltage source supplies energy (causing an electric field and a current), and the resistor converts it to another form (such as thermal energy). In a simple circuit (one with a single simple resistor), the voltage supplied by the source equals the voltage drop across the resistor, since $\text{PE} = q\Delta V$, and the same q flows through each. Thus the energy supplied by the voltage source and the energy converted by the resistor are equal. (See [link].)



The voltage drop across a resistor in a simple circuit equals the voltage output of the battery.

Note:
Making Connections: Conservation of Energy
In a simple electrical circuit, the sole resistor converts energy supplied by the source into another form. Conservation of energy is evidenced here by the fact that all of the energy supplied by the source is converted to another form by the resistor alone. We will find that conservation of energy has other important applications in circuits and is a powerful tool in circuit analysis.

Note:

PhET Explorations: Ohm's Law

See how the equation form of Ohm's law relates to a simple circuit. Adjust the voltage and resistance, and see the current change according to Ohm's law. The sizes of the symbols in the equation change to match the circuit diagram.
https://phet.colorado.edu/sims/html/ohms-law/latest/ohms-law_en.html

Section Summary

- A simple circuit *is* one in which there is a single voltage source and a single resistance.
- One statement of Ohm’s law gives the relationship between current I, voltage V, and resistance R in a simple circuit to be $I=\frac{V}{R}$.
- Resistance has units of ohms (Ω), related to volts and amperes by $1\ \Omega=1\ \text{V}/\text{A}$.
- There is a voltage or IR drop across a resistor, caused by the current flowing through it, given by $V=\text{IR}$.

Conceptual Questions

Exercise:

Problem:

The IR drop across a resistor means that there is a change in potential or voltage across the resistor. Is there any change in current as it passes through a resistor? Explain.

Exercise:

Problem:

How is the IR drop in a resistor similar to the pressure drop in a fluid flowing through a pipe?

Problems & Exercises

Exercise:

Problem:

What current flows through the bulb of a 3.00-V flashlight when its hot resistance is $3.60\ \Omega$?

Solution:

0.833 A

Exercise:

Problem:

Calculate the effective resistance of a pocket calculator that has a 1.35-V battery and through which 0.200 mA flows.

Exercise:

Problem:

What is the effective resistance of a car's starter motor when 150 A flows through it as the car battery applies 11.0 V to the motor?

Solution:

$7.33\times10^{-2}\ \Omega$

Exercise:

Problem:

How many volts are supplied to operate an indicator light on a DVD player that has a resistance of 140 Ω, given that 25.0 mA passes through it?

Exercise:

Problem:

(a) Find the voltage drop in an extension cord having a $0.0600\text{-}\Omega$ resistance and through which 5.00 A is flowing. (b) A cheaper cord utilizes thinner wire and has a resistance of $0.300\ \Omega$. What is the voltage drop in it when 5.00 A flows? (c) Why is the voltage to whatever appliance is being used reduced by this amount? What is the effect on the appliance?

Solution:

(a) 0.300 V

(b) 1.50 V

(c) The voltage supplied to whatever appliance is being used is reduced because the total voltage drop from the wall to the final output of the appliance is fixed. Thus, if the voltage drop across the extension cord is large, the voltage drop across the appliance is significantly decreased, so the power output by the appliance can be significantly decreased, reducing the ability of the appliance to work properly.

Exercise:

Problem:

A power transmission line is hung from metal towers with glass insulators having a resistance of $1.00\times10^{9}\ \Omega$. What current flows through the insulator if the voltage is 200 kV? (Some high-voltage lines are DC.)

Glossary

Ohm's law

an empirical relation stating that the current I is proportional to the potential difference V, $\propto V$; it is often written as $I = V/R$, where R is the resistance

resistance

the electric property that impedes current; for ohmic materials, it is the ratio of voltage to current, $R = V/I$

ohm

the unit of resistance, given by $1\Omega = 1\ \text{V/A}$

ohmic

a type of a material for which Ohm's law is valid

simple circuit

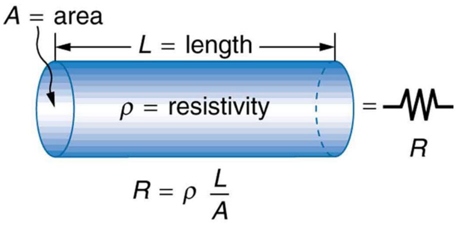
a circuit with a single voltage source and a single resistor

Resistance and Resistivity

- Explain the concept of resistivity.
- Use resistivity to calculate the resistance of specified configurations of material.
- Use the thermal coefficient of resistivity to calculate the change of resistance with temperature.

Material and Shape Dependence of Resistance

The resistance of an object depends on its shape and the material of which it is composed. The cylindrical resistor in [link] is easy to analyze, and, by so doing, we can gain insight into the resistance of more complicated shapes. As you might expect, the cylinder’s electric resistance R is directly proportional to its length L, similar to the resistance of a pipe to fluid flow. The longer the cylinder, the more collisions charges will make with its atoms. The greater the diameter of the cylinder, the more current it can carry (again similar to the flow of fluid through a pipe). In fact, R is inversely proportional to the cylinder’s cross-sectional area A.



A uniform cylinder of length L and cross-sectional area A. Its resistance to the flow of current is similar to the resistance posed by a pipe to fluid flow. The longer the cylinder, the greater its resistance. The larger its cross-sectional area A, the smaller its resistance.

For a given shape, the resistance depends on the material of which the object is composed. Different materials offer different resistance to the flow of charge. We define the **resistivity** ρ of a substance so that the **resistance** R of an object is directly proportional to ρ. Resistivity ρ is an *intrinsic* property of a material, independent of its shape or size. The resistance R of a uniform cylinder of length L, of cross-sectional area A, and made of a material with resistivity ρ, is

Equation:

$$R = \frac{\rho L}{A}.$$

[link] gives representative values of ρ. The materials listed in the table are separated into categories of conductors, semiconductors, and insulators, based on broad groupings of resistivities. Conductors have the smallest resistivities, and insulators have the largest; semiconductors have intermediate resistivities. Conductors have varying but large free charge densities, whereas most charges in insulators are bound to atoms and are not free to move. Semiconductors are intermediate, having far fewer free charges than conductors, but having properties that make the number of free charges depend strongly on the type and amount of impurities in the semiconductor. These unique properties of semiconductors are put to use in modern electronics, as will be explored in later chapters.

Material	**Resistivity ρ ($\Omega \cdot \mathrm{m}$)**
Conductors	
Silver	1.59×10^{-8}
Copper	1.72×10^{-8}
Gold	2.44×10^{-8}
Aluminum	2.65×10^{-8}
Tungsten	5.6×10^{-8}
Iron	9.71×10^{-8}
Platinum	10.6×10^{-8}
Steel	20×10^{-8}
Lead	22×10^{-8}

Material	**Resistivity ρ ($\Omega \cdot \mathrm{m}$)**
Manganin (Cu, Mn, Ni alloy)	44×10^{-8}
Constantan (Cu, Ni alloy)	49×10^{-8}
Mercury	96×10^{-8}
Nichrome (Ni, Fe, Cr alloy)	100×10^{-8}
Semiconductors[footnote] Values depend strongly on amounts and types of impurities	
Carbon (pure)	3.5×10^{-5}
Carbon	$(3.5 - 60) \times 10^{-5}$
Germanium (pure)	600×10^{-3}
Germanium	$(1 - 600) \times 10^{-3}$

Material	**Resistivity ρ ($\Omega \cdot \mathrm{m}$)**
Silicon (pure)	2300
Silicon	$0.1–2300$
Insulators	
Amber	5×10^{14}
Glass	$10^9 - 10^{14}$
Lucite	$>10^{13}$
Mica	$10^{11} - 10^{15}$
Quartz (fused)	75×10^{16}
Rubber (hard)	$10^{13} - 10^{16}$
Sulfur	10^{15}

Material	**Resistivity ρ ($\Omega \cdot \text{m}$)**
Teflon	$>10^{13}$
Wood	$10^{8}-10^{11}$

Resistivities ρ of Various materials at 20^{o}C

Example:

Calculating Resistor Diameter: A Headlight Filament

A car headlight filament is made of tungsten and has a cold resistance of $0.350\ \Omega$. If the filament is a cylinder 4.00 cm long (it may be coiled to save space), what is its diameter?

Strategy

We can rearrange the equation $R=\frac{\rho L}{A}$ to find the cross-sectional area A of the filament from the given information. Then its diameter can be found by assuming it has a circular cross-section.

Solution

The cross-sectional area, found by rearranging the expression for the resistance of a cylinder given in $R=\frac{\rho L}{A}$, is

Equation:

$$A=\frac{\rho L}{R}.$$

Substituting the given values, and taking ρ from [link], yields

Equation:

$$\begin{array}{lll}A & = & \frac{(5.6\times 10^{-8}\ \Omega\cdot\text{m})(4.00\times 10^{-2}\ \text{m})}{0.350\ \Omega}\\ & = & 6.40\times 10^{-9}\ \text{m}^2.\end{array}$$

The area of a circle is related to its diameter D by

Equation:

$$A=\frac{\pi D^2}{4}.$$

Solving for the diameter D, and substituting the value found for A, gives

Equation:

$$\begin{aligned} D &= 2\left(\frac{A}{p}\right)^{\frac{1}{2}}=2\left(\frac{6.40\times10^{-9}\text{ m}^2}{3.14}\right)^{\frac{1}{2}} \\ &= 9.0\times10^{-5}\text{ m}. \end{aligned}$$

Discussion

The diameter is just under a tenth of a millimeter. It is quoted to only two digits, because ρ is known to only two digits.

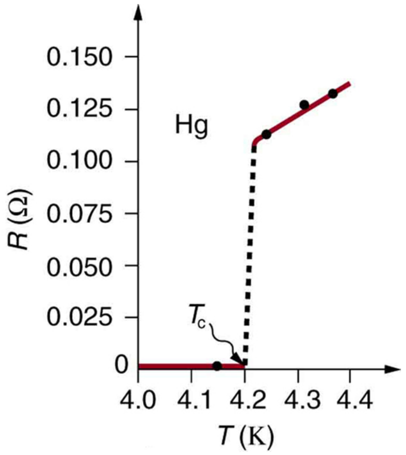
Temperature Variation of Resistance

The resistivity of all materials depends on temperature. Some even become superconductors (zero resistivity) at very low temperatures. (See [link].) Conversely, the resistivity of conductors increases with increasing temperature. Since the atoms vibrate more rapidly and over larger distances at higher temperatures, the electrons moving through a metal make more collisions, effectively making the resistivity higher. Over relatively small temperature changes (about 100ºC or less), resistivity ρ varies with temperature change ΔT as expressed in the following equation

Equation:

$$\rho=\rho_0(1+\alpha\Delta T),$$

where ρ_0 is the original resistivity and α is the **temperature coefficient of resistivity**. (See the values of α in [link] below.) For larger temperature changes, α may vary or a nonlinear equation may be needed to find ρ. Note that α is positive for metals, meaning their resistivity increases with temperature. Some alloys have been developed specifically to have a small temperature dependence. Manganin (which is made of copper, manganese and nickel), for example, has α close to zero (to three digits on the scale in [link]), and so its resistivity varies only slightly with temperature. This is useful for making a temperature-independent resistance standard, for example.



The resistance of a sample of mercury is zero at very low temperatures—it is a superconductor up to about 4.2 K. Above that critical temperature, its resistance makes a sudden jump and then increases nearly linearly with temperature.

Material	**Coefficient α(1/°C)[footnote] Values at 20°C.**
Conductors	
Silver	3.8×10^{-3}
Copper	3.9×10^{-3}
Gold	3.4×10^{-3}
Aluminum	3.9×10^{-3}
Tungsten	4.5×10^{-3}
Iron	5.0×10^{-3}
Platinum	3.93×10^{-3}
Lead	3.9×10^{-3}
Manganin (Cu, Mn, Ni alloy)	0.000×10^{-3}

Material	**Coefficient α(1/°C)[footnote] Values at 20°C.**
Constantan (Cu, Ni alloy)	0.002×10^{-3}
Mercury	0.89×10^{-3}
Nichrome (Ni, Fe, Cr alloy)	0.4×10^{-3}
Semiconductors	
Carbon (pure)	-0.5×10^{-3}
Germanium (pure)	-50×10^{-3}
Silicon (pure)	-70×10^{-3}

Tempature Coefficients of Resistivity α

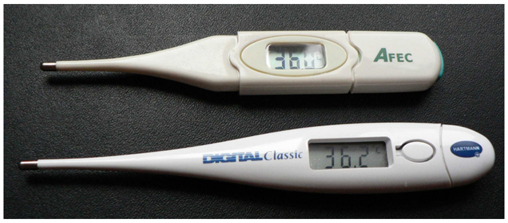
Note also that α is negative for the semiconductors listed in [link], meaning that their resistivity decreases with increasing temperature. They become better conductors at higher temperature, because increased thermal agitation increases the number of free charges available to carry current. This property of decreasing ρ with temperature is also related to the type and amount of impurities present in the semiconductors.

The resistance of an object also depends on temperature, since R_0 is directly proportional to ρ. For a cylinder we know $R = \rho L/A$, and so, if L and A do not change greatly with temperature, R will have the same temperature dependence as ρ. (Examination of the coefficients of linear expansion shows them to be about two orders of magnitude less than typical temperature coefficients of resistivity, and so the effect of temperature on L and A is about two orders of magnitude less than on ρ.) Thus,

Equation:

$$R = R_0(1 + \alpha\Delta T)$$

is the temperature dependence of the resistance of an object, where R_0 is the original resistance and R is the resistance after a temperature change ΔT. Numerous thermometers are based on the effect of temperature on resistance. (See [link].) One of the most common is the thermistor, a semiconductor crystal with a strong temperature dependence, the resistance of which is measured to obtain its temperature. The device is small, so that it quickly comes into thermal equilibrium with the part of a person it touches.



These familiar thermometers are based on the automated measurement of a thermistor’s temperature-dependent resistance. (credit: Biol, Wikimedia Commons)

Example:

Calculating Resistance: Hot-Filament Resistance

Although caution must be used in applying $\rho = \rho_0(1 + \alpha\Delta T)$ and $R = R_0(1 + \alpha\Delta T)$ for temperature changes greater than 100^oC, for tungsten the equations work reasonably well for very large temperature changes. What, then, is the resistance of the tungsten filament in the previous example if its temperature is increased from room temperature (20^oC) to a typical operating temperature of 2850^oC?

Strategy

This is a straightforward application of $R = R_0(1 + \alpha\Delta T)$, since the original resistance of the filament was given to be $R_0 = 0.350\ \Omega$, and the temperature change is $\Delta T = 2830^\text{o}\text{C}$.

Solution

The hot resistance R is obtained by entering known values into the above equation:

Equation:

$$\begin{aligned} R &= R_0(1 + \alpha\Delta T) \\ &= (0.350\ \Omega)[1 + (4.5\times10^{-3}/^\text{o}\text{C})(2830^\text{o}\text{C})] \\ &= 4.8\ \Omega. \end{aligned}$$

Discussion

This value is consistent with the headlight resistance example in Ohm’s Law: Resistance and Simple Circuits.

Note:

PhET Explorations: Resistance in a Wire

Learn about the physics of resistance in a wire. Change its resistivity, length, and area to see how they affect the wire's resistance. The sizes of the symbols in the equation change along with the diagram of a wire.

https://phet.colorado.edu/sims/html/resistance-in-a-wire/latest/resistance-in-a-wire_en.html

Section Summary

- The resistance R of a cylinder of length L and cross-sectional area A is $R=\frac{\rho L}{A}$, where ρ is the resistivity of the material.
- Values of ρ in [link] show that materials fall into three groups—*conductors, semiconductors, and insulators*.
- Temperature affects resistivity; for relatively small temperature changes ΔT, resistivity is $\rho=\rho_0(1+\alpha\Delta T)$, where ρ_0 is the original resistivity and α is the temperature coefficient of resistivity.
- [link] gives values for α, the temperature coefficient of resistivity.
- The resistance R of an object also varies with temperature: $R=R_0(1+\alpha\Delta T)$, where R_0 is the original resistance, and R is the resistance after the temperature change.

Conceptual Questions

Exercise:

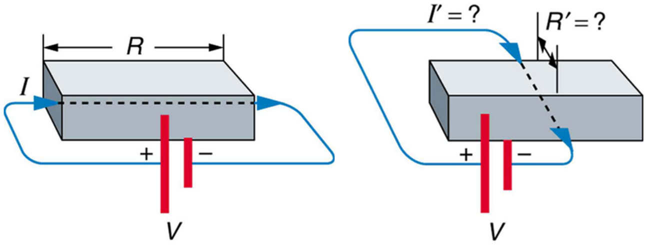
Problem:

In which of the three semiconducting materials listed in [link] do impurities supply free charges? (Hint: Examine the range of resistivity for each and determine whether the pure semiconductor has the higher or lower conductivity.)

Exercise:

Problem:

Does the resistance of an object depend on the path current takes through it? Consider, for example, a rectangular bar—is its resistance the same along its length as across its width? (See [link].)



Does current taking two different paths through the same object encounter different resistance?

Exercise:

Problem:

If aluminum and copper wires of the same length have the same resistance, which has the larger diameter? Why?

Exercise:

Problem:

Explain why $R = R_0(1 + \alpha\Delta T)$ for the temperature variation of the resistance R of an object is not as accurate as $\rho = \rho_0(1 + \alpha\Delta T)$, which gives the temperature variation of resistivity ρ.

Problems & Exercises

Exercise:

Problem:

What is the resistance of a 20.0-m-long piece of 12-gauge copper wire having a 2.053-mm diameter?

Solution:

0.104 Ω

Exercise:

Problem:

The diameter of 0-gauge copper wire is 8.252 mm. Find the resistance of a 1.00-km length of such wire used for power transmission.

Exercise:

Problem:

If the 0.100-mm diameter tungsten filament in a light bulb is to have a resistance of $0.200\ \Omega$ at 20.0°C, how long should it be?

Solution:

$2.8{\times}10^{-2}\text{ m}$

Exercise:

Problem:

Find the ratio of the diameter of aluminum to copper wire, if they have the same resistance per unit length (as they might in household wiring).

Exercise:

Problem:

What current flows through a 2.54-cm-diameter rod of pure silicon that is 20.0 cm long, when $1.00 \times 10^{3}\text{ V}$ is applied to it? (Such a rod may be used to make nuclear-particle detectors, for example.)

Solution:

$1.10{\times}10^{-3}\text{ A}$

Exercise:

Problem:

(a) To what temperature must you raise a copper wire, originally at 20.0ºC, to double its resistance, neglecting any changes in dimensions? (b) Does this happen in household wiring under ordinary circumstances?

Exercise:

Problem:

A resistor made of Nichrome wire is used in an application where its resistance cannot change more than 1.00% from its value at 20.0ºC. Over what temperature range can it be used?

Solution:

-5°C to 45°C

Exercise:

Problem:

Of what material is a resistor made if its resistance is 40.0% greater at 100ºC than at 20.0ºC?

Exercise:

Problem:

An electronic device designed to operate at any temperature in the range from –10.0ºC to 55.0ºC contains pure carbon resistors. By what factor does their resistance increase over this range?

Solution:

1.03

Exercise:

Problem:

(a) Of what material is a wire made, if it is 25.0 m long with a 0.100 mm diameter and has a resistance of $77.7\ \Omega$ at 20.0°C? (b) What is its resistance at 150°C?

Exercise:

Problem:

Assuming a constant temperature coefficient of resistivity, what is the maximum percent decrease in the resistance of a constantan wire starting at 20.0°C?

Solution:

0.06%

Exercise:

Problem:

A wire is drawn through a die, stretching it to four times its original length. By what factor does its resistance increase?

Exercise:

Problem:

A copper wire has a resistance of $0.500\ \Omega$ at 20.0°C, and an iron wire has a resistance of $0.525\ \Omega$ at the same temperature. At what temperature are their resistances equal?

Solution:

-17°C

Exercise:

Problem:

(a) Digital medical thermometers determine temperature by measuring the resistance of a semiconductor device called a thermistor (which has $\alpha = -0.0600/^{\circ}\text{C}$) when it is at the same temperature as the patient. What is a patient's temperature if the thermistor's resistance at that temperature is 82.0% of its value at 37.0°C (normal body temperature)? (b) The negative value for α may not be maintained for very low temperatures. Discuss why and whether this is the case here. (Hint: Resistance can't become negative.)

Exercise:

Problem: Integrated Concepts

(a) Redo [link] taking into account the thermal expansion of the tungsten filament. You may assume a thermal expansion coefficient of $12 \times 10^{-6}/^{\circ}\text{C}$. (b) By what percentage does your answer differ from that in the example?

Solution:

(a) $4.7\ \Omega$ (total)

(b) 3.0% decrease

Exercise:

Problem: Unreasonable Results

(a) To what temperature must you raise a resistor made of constantan to double its resistance, assuming a constant temperature coefficient of resistivity? (b) To cut it in half? (c) What is unreasonable about these results? (d) Which assumptions are unreasonable, or which premises are inconsistent?

Glossary

resistivity

an intrinsic property of a material, independent of its shape or size, directly proportional to the resistance, denoted by ρ

temperature coefficient of resistivity

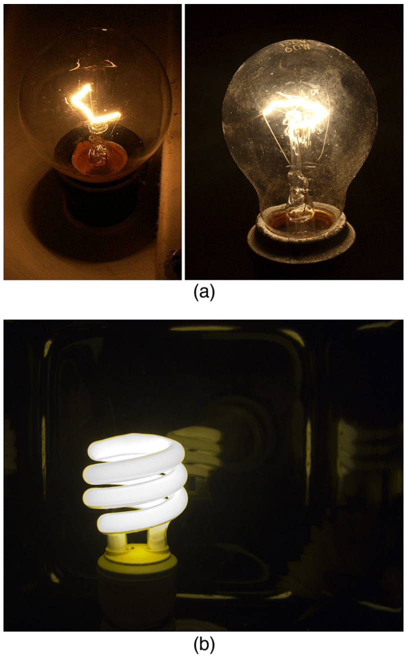
an empirical quantity, denoted by α, which describes the change in resistance or resistivity of a material with temperature

Electric Power and Energy

- Calculate the power dissipated by a resistor and power supplied by a power supply.
- Calculate the cost of electricity under various circumstances.

Power in Electric Circuits

Power is associated by many people with electricity. Knowing that power is the rate of energy use or energy conversion, what is the expression for **electric power**? Power transmission lines might come to mind. We also think of lightbulbs in terms of their power ratings in watts. Let us compare a 25-W bulb with a 60-W bulb. (See [link](a).) Since both operate on the same voltage, the 60-W bulb must draw more current to have a greater power rating. Thus the 60-W bulb's resistance must be lower than that of a 25-W bulb. If we increase voltage, we also increase power. For example, when a 25-W bulb that is designed to operate on 120 V is connected to 240 V, it briefly glows very brightly and then burns out. Precisely how are voltage, current, and resistance related to electric power?



(a) Which of these lightbulbs, the 25-W bulb (upper left) or the 60-W bulb (upper right), has the higher resistance? Which draws more current? Which uses the most energy? Can you tell from the color that the 25-W filament is cooler? Is the brighter bulb a different color and if so why? (credits: Dickbauch, Wikimedia Commons; Greg Westfall, Flickr) (b) This compact fluorescent light (CFL) puts out the same intensity of light as the 60-W bulb, but at 1/4 to 1/10 the input power. (credit: dbgg1979, Flickr)

Electric energy depends on both the voltage involved and the charge moved. This is expressed most simply as $\text{PE} = \text{qV}$, where q is the charge moved and V is the voltage (or more precisely, the potential difference the charge moves through). Power is the rate at which energy is moved, and so electric power is

Equation:

$$P=\frac{\text{PE}}{t}=\frac{\text{qV}}{t}.$$

Recognizing that current is $I=q/t$ (note that $\Delta t=t$ here), the expression for power becomes

Equation:

$$P=\text{IV}.$$

Electric power (P) is simply the product of current times voltage. Power has familiar units of watts. Since the SI unit for potential energy (PE) is the joule, power has units of joules per second, or watts. Thus, $1\text{ A}\cdot\text{V}=1\text{ W}$. For example, cars often have one or more auxiliary power outlets with which you can charge a cell phone or other electronic devices. These outlets may be rated at 20 A, so that the circuit can deliver a maximum power $P=\text{IV}=(20\text{ A})(12\text{ V})=240\text{ W}$. In some applications, electric power may be expressed as volt-amperes or even kilovolt-amperes ($1\text{ kA}\cdot\text{V}=1\text{ kW}$).

To see the relationship of power to resistance, we combine Ohm’s law with $P=\text{IV}$. Substituting $I=\text{V/R}$ gives $P=(V/R)V=V^2/R$. Similarly, substituting $V=\text{IR}$ gives $P=I(\text{IR})=I^2R$. Three expressions for electric power are listed together here for convenience:

Equation:

$$P=\text{IV}$$

Equation:

$$P=\frac{V^2}{R}$$

Equation:

$$P = I^2 R.$$

Note that the first equation is always valid, whereas the other two can be used only for resistors. In a simple circuit, with one voltage source and a single resistor, the power supplied by the voltage source and that dissipated by the resistor are identical. (In more complicated circuits, P can be the power dissipated by a single device and not the total power in the circuit.)

Different insights can be gained from the three different expressions for electric power. For example, $P = V^2/R$ implies that the lower the resistance connected to a given voltage source, the greater the power delivered. Furthermore, since voltage is squared in $P = V^2/R$, the effect of applying a higher voltage is perhaps greater than expected. Thus, when the voltage is doubled to a 25-W bulb, its power nearly quadruples to about 100 W, burning it out. If the bulb's resistance remained constant, its power would be exactly 100 W, but at the higher temperature its resistance is higher, too.

Example:
Calculating Power Dissipation and Current: Hot and Cold Power
(a) Consider the examples given in Ohm's Law: Resistance and Simple Circuits and Resistance and Resistivity. Then find the power dissipated by the car headlight in these examples, both when it is hot and when it is cold. (b) What current does it draw when cold?

Strategy for (a)

For the hot headlight, we know voltage and current, so we can use $P = \mathrm{IV}$ to find the power. For the cold headlight, we know the voltage and resistance, so we can use $P = V^2/R$ to find the power.

Solution for (a)

Entering the known values of current and voltage for the hot headlight, we obtain

Equation:

$$P = \mathrm{IV} = (2.50\ \mathrm{A})(12.0\ \mathrm{V}) = 30.0\ \mathrm{W}.$$

The cold resistance was $0.350\ \Omega$, and so the power it uses when first switched on is

Equation:

$$P=\frac{V^2}{R}=\frac{(12.0\text{ V})^2}{0.350\ \Omega}=411\text{ W}.$$

Discussion for (a)

The 30 W dissipated by the hot headlight is typical. But the 411 W when cold is surprisingly higher. The initial power quickly decreases as the bulb’s temperature increases and its resistance increases.

Strategy and Solution for (b)

The current when the bulb is cold can be found several different ways. We rearrange one of the power equations, $P=I^2R$, and enter known values, obtaining

Equation:

$$I=\sqrt{\frac{P}{R}}=\sqrt{\frac{411\text{ W}}{0.350\ \Omega}}=34.3\text{ A}.$$

Discussion for (b)

The cold current is remarkably higher than the steady-state value of 2.50 A, but the current will quickly decline to that value as the bulb’s temperature increases. Most fuses and circuit breakers (used to limit the current in a circuit) are designed to tolerate very high currents briefly as a device comes on. In some cases, such as with electric motors, the current remains high for several seconds, necessitating special “slow blow” fuses.

The Cost of Electricity

The more electric appliances you use and the longer they are left on, the higher your electric bill. This familiar fact is based on the relationship between energy and power. You pay for the energy used. Since $P=E/t$, we see that

Equation:

$$E = \text{Pt}$$

is the energy used by a device using power P for a time interval t. For example, the more lightbulbs burning, the greater P used; the longer they are on, the greater t is. The energy unit on electric bills is the kilowatt-hour $(\text{kW}\cdot\text{h})$, consistent with the relationship $E = \text{Pt}$. It is easy to estimate the cost of operating electric appliances if you have some idea of their power consumption rate in watts or kilowatts, the time they are on in hours, and the cost per kilowatt-hour for your electric utility. Kilowatt-hours, like all other specialized energy units such as food calories, can be converted to joules. You can prove to yourself that $1\ \text{kW}\cdot\text{h} = 3.6{\times}10^{6}\ \text{J}$.

The electrical energy (E) used can be reduced either by reducing the time of use or by reducing the power consumption of that appliance or fixture. This will not only reduce the cost, but it will also result in a reduced impact on the environment. Improvements to lighting are some of the fastest ways to reduce the electrical energy used in a home or business. About 20% of a home's use of energy goes to lighting, while the number for commercial establishments is closer to 40%. Fluorescent lights are about four times more efficient than incandescent lights—this is true for both the long tubes and the compact fluorescent lights (CFL). (See [link](b).) Thus, a 60-W incandescent bulb can be replaced by a 15-W CFL, which has the same brightness and color. CFLs have a bent tube inside a globe or a spiral-shaped tube, all connected to a standard screw-in base that fits standard incandescent light sockets. (Original problems with color, flicker, shape, and high initial investment for CFLs have been addressed in recent years.) The heat transfer from these CFLs is less, and they last up to 10 times longer. The significance of an investment in such bulbs is addressed in the next example. New white LED lights (which are clusters of small LED bulbs) are even more efficient (twice that of CFLs) and last 5 times longer than CFLs. However, their cost is still high.

Note:
Making Connections: Energy, Power, and Time

The relationship $E = \text{Pt}$ is one that you will find useful in many different contexts. The energy your body uses in exercise is related to the power level and duration of your activity, for example. The amount of heating by a power source is related to the power level and time it is applied. Even the radiation dose of an X-ray image is related to the power and time of exposure.

Example:

Calculating the Cost Effectiveness of Compact Fluorescent Lights (CFL)

If the cost of electricity in your area is 12 cents per kWh, what is the total cost (capital plus operation) of using a 60-W incandescent bulb for 1000 hours (the lifetime of that bulb) if the bulb cost 25 cents? (b) If we replace this bulb with a compact fluorescent light that provides the same light output, but at one-quarter the wattage, and which costs $1.50 but lasts 10 times longer (10,000 hours), what will that total cost be?

Strategy

To find the operating cost, we first find the energy used in kilowatt-hours and then multiply by the cost per kilowatt-hour.

Solution for (a)

The energy used in kilowatt-hours is found by entering the power and time into the expression for energy:

Equation:

$$E = \text{Pt} = (60\text{ W})(1000\text{ h}) = 60{,}000\text{ W}\cdot\text{h}.$$

In kilowatt-hours, this is

Equation:

$$E = 60.0\text{ kW}\cdot\text{h}.$$

Now the electricity cost is

Equation:

$$\text{cost} = (60.0\text{ kW}\cdot\text{h})(\$0.12/\text{kW}\cdot\text{h}) = \$7.20.$$

The total cost will be $7.20 for 1000 hours (about one-half year at 5 hours per day).

Solution for (b)

Since the CFL uses only 15 W and not 60 W, the electricity cost will be $7.20/4 = $1.80. The CFL will last 10 times longer than the incandescent, so that the investment cost will be 1/10 of the bulb cost for that time period of use, or 0.1($1.50) = $0.15. Therefore, the total cost will be $1.95 for 1000 hours.

Discussion

Therefore, it is much cheaper to use the CFLs, even though the initial investment is higher. The increased cost of labor that a business must include for replacing the incandescent bulbs more often has not been figured in here.

Note:

Making Connections: Take-Home Experiment—Electrical Energy Use Inventory

1) Make a list of the power ratings on a range of appliances in your home or room. Explain why something like a toaster has a higher rating than a digital clock. Estimate the energy consumed by these appliances in an average day (by estimating their time of use). Some appliances might only state the operating current. If the household voltage is 120 V, then use $P = \mathrm{IV}$. 2) Check out the total wattage used in the rest rooms of your school's floor or building. (You might need to assume the long fluorescent lights in use are rated at 32 W.) Suppose that the building was closed all weekend and that these lights were left on from 6 p.m. Friday until 8 a.m. Monday. What would this oversight cost? How about for an entire year of weekends?

Section Summary

- Electric power P is the rate (in watts) that energy is supplied by a source or dissipated by a device.

- Three expressions for electrical power are
  Equation:

  $$P = \text{IV},$$

  Equation:

  $$P = \frac{V^2}{R},$$

  and
  Equation:

  $$P = I^2R.$$

- The energy used by a device with a power P over a time t is $E = \text{Pt}$.

Conceptual Questions

Exercise:

Problem:

Why do incandescent lightbulbs grow dim late in their lives, particularly just before their filaments break?

Exercise:

Problem:

The power dissipated in a resistor is given by $P = V^2/R$, which means power decreases if resistance increases. Yet this power is also given by $P = I^2R$, which means power increases if resistance increases. Explain why there is no contradiction here.

Problem Exercises

Exercise:

Problem:

What is the power of a 1.00×10^2 MV lightning bolt having a current of 2.00×10^4 A?

Solution:

$2.00{\times}10^{12}$ W

Exercise:

Problem:

What power is supplied to the starter motor of a large truck that draws 250 A of current from a 24.0-V battery hookup?

Exercise:

Problem:

A charge of 4.00 C of charge passes through a pocket calculator’s solar cells in 4.00 h. What is the power output, given the calculator’s voltage output is 3.00 V? (See [link].)

The strip of solar cells just above the keys of this calculator convert

light to electricity
to supply its energy
needs. (credit:
Evan-Amos,
Wikimedia
Commons)

Exercise:

Problem:

How many watts does a flashlight that has 6.00×10^2 C pass through it in 0.500 h use if its voltage is 3.00 V?

Exercise:

Problem:

Find the power dissipated in each of these extension cords: (a) an extension cord having a $0.0600 - \Omega$ resistance and through which 5.00 A is flowing; (b) a cheaper cord utilizing thinner wire and with a resistance of $0.300\ \Omega$.

Solution:

(a) 1.50 W

(b) 7.50 W

Exercise:

Problem:

Verify that the units of a volt-ampere are watts, as implied by the equation $P = \text{IV}$.

Exercise:

Problem:

Show that the units $1\ \mathrm{V}^2/\Omega = 1\mathrm{W}$, as implied by the equation $P = V^2/R$.

Solution:

$$\frac{V^2}{\Omega} = \frac{V^2}{\mathrm{V/A}} = \mathrm{AV} = \left(\frac{C}{s}\right)\left(\frac{J}{C}\right) = \frac{J}{s} = 1\ \mathrm{W}$$

Exercise:

Problem:

Show that the units $1\ \mathrm{A}^2 \cdot \Omega = 1\ \mathrm{W}$, as implied by the equation $P = I^2R$.

Exercise:

Problem:

Verify the energy unit equivalence that $1\ \mathrm{kW} \cdot \mathrm{h} = 3.60{\times}10^6\ \mathrm{J}$.

Solution:

$$1\ \mathrm{kW} \cdot \mathrm{h}{=}\left(\frac{1{\times}10^3\ \mathrm{J}}{1\ \mathrm{s}}\right)(1\ \mathrm{h})\left(\frac{3600\ \mathrm{s}}{1\ \mathrm{h}}\right) = 3.60 \times 10^6\ \mathrm{J}$$

Exercise:

Problem:

Electrons in an X-ray tube are accelerated through $1.00 \times 10^2\ \mathrm{kV}$ and directed toward a target to produce X-rays. Calculate the power of the electron beam in this tube if it has a current of 15.0 mA.

Exercise:

Problem:

An electric water heater consumes 5.00 kW for 2.00 h per day. What is the cost of running it for one year if electricity costs $12.0\ \text{cents/kW}\cdot\text{h}$? See [link].

On-demand electric hot water heater. Heat is supplied to water only when needed. (credit: aviddavid, Flickr)

Solution:

$438/y

Exercise:

Problem:

With a 1200-W toaster, how much electrical energy is needed to make a slice of toast (cooking time = 1 minute)? At $9.0\ \text{cents/kW}\cdot\text{h}$, how much does this cost?

Exercise:

Problem:

What would be the maximum cost of a CFL such that the total cost (investment plus operating) would be the same for both CFL and incandescent 60-W bulbs? Assume the cost of the incandescent bulb is 25 cents and that electricity costs 10 cents/kWh. Calculate the cost for 1000 hours, as in the cost effectiveness of CFL example.

Solution:

$6.25

Exercise:

Problem:

Some makes of older cars have 6.00-V electrical systems. (a) What is the hot resistance of a 30.0-W headlight in such a car? (b) What current flows through it?

Exercise:

Problem:

Alkaline batteries have the advantage of putting out constant voltage until very nearly the end of their life. How long will an alkaline battery rated at $1.00\text{ A}\cdot\text{h}$ and 1.58 V keep a 1.00-W flashlight bulb burning?

Solution:

1.58 h

Exercise:

Problem:

A cauterizer, used to stop bleeding in surgery, puts out 2.00 mA at 15.0 kV. (a) What is its power output? (b) What is the resistance of the path?

Exercise:

Problem:

The average television is said to be on 6 hours per day. Estimate the yearly cost of electricity to operate 100 million TVs, assuming their power consumption averages 150 W and the cost of electricity averages $12.0\text{ cents/kW}\cdot\text{h}$.

Solution:

$3.94 billion/year

Exercise:

Problem:

An old lightbulb draws only 50.0 W, rather than its original 60.0 W, due to evaporative thinning of its filament. By what factor is its diameter reduced, assuming uniform thinning along its length? Neglect any effects caused by temperature differences.

Exercise:

Problem:

00-gauge copper wire has a diameter of 9.266 mm. Calculate the power loss in a kilometer of such wire when it carries $1.00\times 10^{2}\text{ A}$.

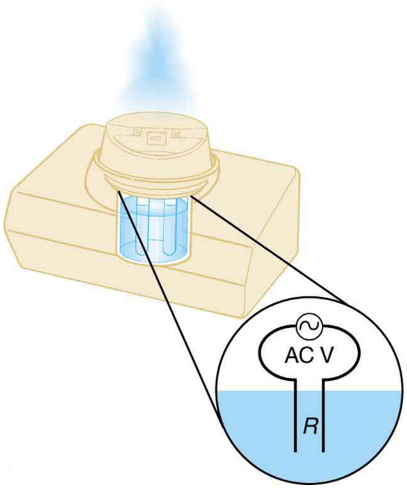
Solution:

25.5 W

Exercise:

Problem: **Integrated Concepts**

Cold vaporizers pass a current through water, evaporating it with only a small increase in temperature. One such home device is rated at 3.50 A and utilizes 120 V AC with 95.0% efficiency. (a) What is the vaporization rate in grams per minute? (b) How much water must you put into the vaporizer for 8.00 h of overnight operation? (See [link].)



This cold vaporizer passes current directly through water, vaporizing it directly with relatively little temperature increase.

Exercise:

Problem: Integrated Concepts

(a) What energy is dissipated by a lightning bolt having a 20,000-A current, a voltage of 1.00×10^2 MV, and a length of 1.00 ms? (b) What mass of tree sap could be raised from 18.0^{o}C to its boiling point and then evaporated by this energy, assuming sap has the same thermal characteristics as water?

Solution:

(a) 2.00×10^9 J

(b) 769 kg

Exercise:

Problem: Integrated Concepts

What current must be produced by a 12.0-V battery-operated bottle warmer in order to heat 75.0 g of glass, 250 g of baby formula, and 3.00×10^2 g of aluminum from 20.0°C to 90.0°C in 5.00 min?

Exercise:

Problem: Integrated Concepts

How much time is needed for a surgical cauterizer to raise the temperature of 1.00 g of tissue from 37.0°C to 100°C and then boil away 0.500 g of water, if it puts out 2.00 mA at 15.0 kV? Ignore heat transfer to the surroundings.

Solution:

45.0 s

Exercise:

Problem: Integrated Concepts

Hydroelectric generators (see [link]) at Hoover Dam produce a maximum current of $8.00 \times 10^3\ \text{A}$ at 250 kV. (a) What is the power output? (b) The water that powers the generators enters and leaves the system at low speed (thus its kinetic energy does not change) but loses 160 m in altitude. How many cubic meters per second are needed, assuming 85.0% efficiency?

Hydroelectric generators at the Hoover dam. (credit: Jon Sullivan)

Exercise:

Problem: Integrated Concepts

(a) Assuming 95.0% efficiency for the conversion of electrical power by the motor, what current must the 12.0-V batteries of a 750-kg electric car be able to supply: (a) To accelerate from rest to 25.0 m/s in 1.00 min? (b) To climb a 2.00×10^{2}-m-high hill in 2.00 min at a constant 25.0-m/s speed while exerting 5.00×10^{2} N of force to overcome air resistance and friction? (c) To travel at a constant 25.0-m/s speed, exerting a 5.00×10^{2} N force to overcome air resistance and friction? See [link].

This REVAi, an electric

car, gets recharged on a street in London. (credit: Frank Hebbert)

Solution:

(a) 343 A

(b) 2.17×10^{3} A

(c) 1.10×10^{3} A

Exercise:

Problem: **Integrated Concepts**

A light-rail commuter train draws 630 A of 650-V DC electricity when accelerating. (a) What is its power consumption rate in kilowatts? (b) How long does it take to reach 20.0 m/s starting from rest if its loaded mass is 5.30×10^{4} kg, assuming 95.0% efficiency and constant power? (c) Find its average acceleration. (d) Discuss how the acceleration you found for the light-rail train compares to what might be typical for an automobile.

Exercise:

Problem: **Integrated Concepts**

(a) An aluminum power transmission line has a resistance of $0.0580\ \Omega/\text{km}$. What is its mass per kilometer? (b) What is the mass per kilometer of a copper line having the same resistance? A lower resistance would shorten the heating time. Discuss the practical limits to speeding the heating by lowering the resistance.

Solution:

(a) 1.23×10^3 kg

(b) 2.64×10^3 kg

Exercise:

Problem: Integrated Concepts

(a) An immersion heater utilizing 120 V can raise the temperature of a 1.00×10^2-g aluminum cup containing 350 g of water from $20.0^\circ\mathrm{C}$ to $95.0^\circ\mathrm{C}$ in 2.00 min. Find its resistance, assuming it is constant during the process. (b) A lower resistance would shorten the heating time. Discuss the practical limits to speeding the heating by lowering the resistance.

Exercise:

Problem: Integrated Concepts

(a) What is the cost of heating a hot tub containing 1500 kg of water from $10.0^\circ\mathrm{C}$ to $40.0^\circ\mathrm{C}$, assuming 75.0% efficiency to account for heat transfer to the surroundings? The cost of electricity is $9\ \mathrm{cents/kW \cdot h}$. (b) What current was used by the 220-V AC electric heater, if this took 4.00 h?

Exercise:

Problem: Unreasonable Results

(a) What current is needed to transmit $1.00 \times 10^2\ \mathrm{MW}$ of power at 480 V? (b) What power is dissipated by the transmission lines if they have a 1.00 - Ω resistance? (c) What is unreasonable about this result? (d) Which assumptions are unreasonable, or which premises are inconsistent?

Solution:

(a) $2.08 \times 10^5\ \mathrm{A}$

(b) 4.33×10^4 MW

(c) The transmission lines dissipate more power than they are supposed to transmit.

(d) A voltage of 480 V is unreasonably low for a transmission voltage. Long-distance transmission lines are kept at much higher voltages (often hundreds of kilovolts) to reduce power losses.

Exercise:

Problem: Unreasonable Results

(a) What current is needed to transmit 1.00×10^2 MW of power at 10.0 kV? (b) Find the resistance of 1.00 km of wire that would cause a 0.0100% power loss. (c) What is the diameter of a 1.00-km-long copper wire having this resistance? (d) What is unreasonable about these results? (e) Which assumptions are unreasonable, or which premises are inconsistent?

Exercise:

Problem: Construct Your Own Problem

Consider an electric immersion heater used to heat a cup of water to make tea. Construct a problem in which you calculate the needed resistance of the heater so that it increases the temperature of the water and cup in a reasonable amount of time. Also calculate the cost of the electrical energy used in your process. Among the things to be considered are the voltage used, the masses and heat capacities involved, heat losses, and the time over which the heating takes place. Your instructor may wish for you to consider a thermal safety switch (perhaps bimetallic) that will halt the process before damaging temperatures are reached in the immersion unit.

Glossary

electric power

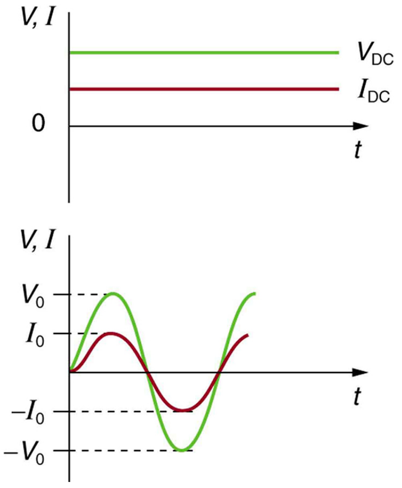
the rate at which electrical energy is supplied by a source or dissipated by a device; it is the product of current times voltage

Alternating Current versus Direct Current

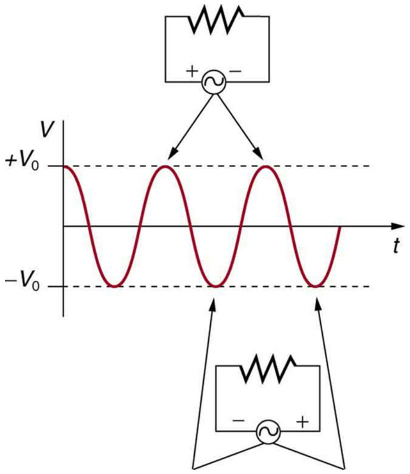
- Explain the differences and similarities between AC and DC current.
- Calculate rms voltage, current, and average power.
- Explain why AC current is used for power transmission.

Alternating Current

Most of the examples dealt with so far, and particularly those utilizing batteries, have constant voltage sources. Once the current is established, it is thus also a constant. **Direct current** (DC) is the flow of electric charge in only one direction. It is the steady state of a constant-voltage circuit. Most well-known applications, however, use a time-varying voltage source. **Alternating current** (AC) is the flow of electric charge that periodically reverses direction. If the source varies periodically, particularly sinusoidally, the circuit is known as an alternating current circuit. Examples include the commercial and residential power that serves so many of our needs. [link] shows graphs of voltage and current versus time for typical DC and AC power. The AC voltages and frequencies commonly used in homes and businesses vary around the world.



(a) DC voltage and current are constant in time, once the current is established. (b) A graph of voltage and current versus time for 60-Hz AC power. The voltage and current are sinusoidal and are in phase for a simple resistance circuit. The frequencies and peak voltages of AC sources differ greatly.



The potential difference V between the terminals of an AC voltage source fluctuates as

shown. The
mathematical
expression for V is
given by
$V = V_0 \sin 2\pi\text{ft}$.

[link] shows a schematic of a simple circuit with an AC voltage source. The voltage between the terminals fluctuates as shown, with the **AC voltage** given by

Equation:

$$V = V_0 \sin 2\pi\text{ft},$$

where V is the voltage at time t, V_0 is the peak voltage, and f is the frequency in hertz. For this simple resistance circuit, $I = \text{V}/\text{R}$, and so the **AC current** is

Equation:

$$I = I_0 \sin 2\pi\text{ft},$$

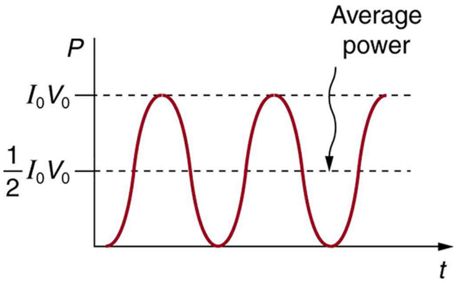
where I is the current at time t, and $I_0 = V_0/\text{R}$ is the peak current. For this example, the voltage and current are said to be in phase, as seen in [link](b).

Current in the resistor alternates back and forth just like the driving voltage, since $I = \text{V}/\text{R}$. If the resistor is a fluorescent light bulb, for example, it brightens and dims 120 times per second as the current repeatedly goes through zero. A 120-Hz flicker is too rapid for your eyes to detect, but if you wave your hand back and forth between your face and a fluorescent light, you will see a stroboscopic effect evidencing AC. The fact that the light output fluctuates means that the power is fluctuating. The power supplied is $P = \text{IV}$. Using the expressions for I and V above, we see that the time dependence of power is $P = I_0 V_0 \sin^2 2\pi\text{ft}$, as shown in [link].

Note:

Making Connections: Take-Home Experiment—AC/DC Lights

Wave your hand back and forth between your face and a fluorescent light bulb. Do you observe the same thing with the headlights on your car? Explain what you observe. *Warning: Do not look directly at very bright light.*



AC power as a function of time. Since the voltage and current are in phase here, their product is non-negative and fluctuates between zero and I_0V_0. Average power is $(1/2)I_0V_0$.

We are most often concerned with average power rather than its fluctuations—that 60-W light bulb in your desk lamp has an average power consumption of 60 W, for example. As illustrated in [link], the average power P_{ave} is

Equation:

$$P_{\text{ave}} = \frac{1}{2} I_0 V_0.$$

This is evident from the graph, since the areas above and below the $(1/2)I_0V_0$ line are equal, but it can also be proven using trigonometric identities. Similarly, we define an average or **rms current** I_{rms} and average or **rms voltage** V_{rms} to be, respectively,

Equation:

$$I_{\text{rms}} = \frac{I_0}{\sqrt{2}}$$

and

Equation:

$$V_{\text{rms}} = \frac{V_0}{\sqrt{2}}.$$

where rms stands for root mean square, a particular kind of average. In general, to obtain a root mean square, the particular quantity is squared, its mean (or average) is found, and the square root is taken. This is useful for AC, since the average value is zero. Now,

Equation:

$$P_{\text{ave}} = I_{\text{rms}}V_{\text{rms}},$$

which gives

Equation:

$$P_{\text{ave}} = \frac{I_0}{\sqrt{2}} \cdot \frac{V_0}{\sqrt{2}} = \frac{1}{2}I_0V_0,$$

as stated above. It is standard practice to quote I_{rms}, V_{rms}, and P_{ave} rather than the peak values. For example, most household electricity is 120 V AC, which means that V_{rms} is 120 V. The common 10-A circuit breaker will interrupt a sustained I_{rms} greater than 10 A. Your 1.0-kW microwave oven consumes $P_{\text{ave}} = 1.0\text{ kW}$, and so on. You can think of these rms and average values as the equivalent DC values for a simple resistive circuit.

To summarize, when dealing with AC, Ohm’s law and the equations for power are completely analogous to those for DC, but rms and average values are used for AC. Thus, for AC, Ohm’s law is written

Equation:

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{R}.$$

The various expressions for AC power P_{ave} are

Equation:

$$P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}},$$

Equation:

$$P_{\text{ave}} = \frac{V_{\text{rms}}^2}{R},$$

and

Equation:

$$P_{\text{ave}} = I_{\text{rms}}^2 R.$$

Example:

Peak Voltage and Power for AC

(a) What is the value of the peak voltage for 120-V AC power? (b) What is the peak power consumption rate of a 60.0-W AC light bulb?

Strategy

We are told that V_{rms} is 120 V and P_{ave} is 60.0 W. We can use $V_{\text{rms}} = \frac{V_0}{\sqrt{2}}$ to find the peak voltage, and we can manipulate the definition of power to find the peak power from the given average power.

Solution for (a)

Solving the equation $V_{\text{rms}} = \frac{V_0}{\sqrt{2}}$ for the peak voltage V_0 and substituting the known value for V_{rms} gives

Equation:

$$V_0 = \sqrt{2}V_{\text{rms}} = 1.414(120\ \text{V}) = 170\ \text{V}.$$

Discussion for (a)

This means that the AC voltage swings from 170 V to –170 V and back 60 times every second. An equivalent DC voltage is a constant 120 V.

Solution for (b)

Peak power is peak current times peak voltage. Thus,

Equation:

$$P_0 = I_0V_0 = 2\left(\frac{1}{2}I_0V_0\right) = 2P_{\text{ave}}.$$

We know the average power is 60.0 W, and so

Equation:

$$P_0 = 2(60.0\ \text{W}) = 120\ \text{W}.$$

Discussion

So the power swings from zero to 120 W one hundred twenty times per second (twice each cycle), and the power averages 60 W.

Why Use AC for Power Distribution?

Most large power-distribution systems are AC. Moreover, the power is transmitted at much higher voltages than the 120-V AC (240 V in most parts of the world) we use in homes and on the job. Economies of scale make it cheaper to build a few very large electric power-generation plants than to build numerous small ones. This necessitates sending power long distances, and it is obviously important that energy losses en route be minimized. High voltages can be transmitted with much smaller power losses than low voltages, as we shall see. (See [link].) For safety reasons, the voltage at the user is reduced to familiar values. The crucial factor is that it is much easier to increase and decrease AC voltages than DC, so AC is used in most large power distribution systems.

Power is distributed over large distances at high voltage to reduce power loss in the transmission lines. The voltages generated at the power plant are stepped up by passive devices called transformers (see Transformers) to 330,000 volts (or more in some places worldwide). At the point of use, the transformers reduce the voltage transmitted for safe residential and commercial use. (Credit: GeorgHH, Wikimedia Commons)

Example:

Power Losses Are Less for High-Voltage Transmission

(a) What current is needed to transmit 100 MW of power at 200 kV? (b) What is the power dissipated by the transmission lines if they have a resistance of 1.00 Ω? (c) What percentage of the power is lost in the transmission lines?

Strategy

We are given $P_{\text{ave}} = 100\text{ MW}$, $V_{\text{rms}} = 200\text{ kV}$, and the resistance of the lines is $R = 1.00\ \Omega$. Using these givens, we can find the current flowing (from $P = \text{IV}$) and then the power dissipated in the lines ($P = I^2R$), and we take the ratio to the total power transmitted.

Solution

To find the current, we rearrange the relationship $P_{\text{ave}} = I_{\text{rms}}V_{\text{rms}}$ and substitute known values. This gives

Equation:

$$I_{\text{rms}} = \frac{P_{\text{ave}}}{V_{\text{rms}}} = \frac{100 \times 10^6\text{ W}}{200 \times 10^3\text{ V}} = 500\text{ A}.$$

Solution

Knowing the current and given the resistance of the lines, the power dissipated in them is found from $P_{\text{ave}} = I_{\text{rms}}^2R$. Substituting the known values gives

Equation:

$$P_{\text{ave}} = I_{\text{rms}}^2R = (500\text{ A})^2(1.00\ \Omega) = 250\text{ kW}.$$

Solution

The percent loss is the ratio of this lost power to the total or input power, multiplied by 100:

Equation:

$$\% \text{ loss} = \frac{250 \text{ kW}}{100 \text{ MW}} \times 100 = 0.250 \text{ \%}.$$

Discussion

One-fourth of a percent is an acceptable loss. Note that if 100 MW of power had been transmitted at 25 kV, then a current of 4000 A would have been needed. This would result in a power loss in the lines of 16.0 MW, or 16.0% rather than 0.250%. The lower the voltage, the more current is needed, and the greater the power loss in the fixed-resistance transmission lines. Of course, lower-resistance lines can be built, but this requires larger and more expensive wires. If superconducting lines could be economically produced, there would be no loss in the transmission lines at all. But, as we shall see in a later chapter, there is a limit to current in superconductors, too. In short, high voltages are more economical for transmitting power, and AC voltage is much easier to raise and lower, so that AC is used in most large-scale power distribution systems.

It is widely recognized that high voltages pose greater hazards than low voltages. But, in fact, some high voltages, such as those associated with common static electricity, can be harmless. So it is not voltage alone that determines a hazard. It is not so widely recognized that AC shocks are often more harmful than similar DC shocks. Thomas Edison thought that AC shocks were more harmful and set up a DC power-distribution system in New York City in the late 1800s. There were bitter fights, in particular between Edison and George Westinghouse and Nikola Tesla, who were advocating the use of AC in early power-distribution systems. AC has prevailed largely due to transformers and lower power losses with high-voltage transmission.

Note:

PhET Explorations: Generator

Generate electricity with a bar magnet! Discover the physics behind the phenomena by exploring magnets and how you can use them to make a bulb light.

Generator

Section Summary

- Direct current (DC) is the flow of electric current in only one direction. It refers to systems where the source voltage is constant.
- The voltage source of an alternating current (AC) system puts out $V = V_0 \text{ sin } 2\pi\text{ft}$, where V is the voltage at time t, V_0 is the peak voltage, and f is the frequency in hertz.
- In a simple circuit, $I = \text{V}/\text{R}$ and AC current is $I = I_0 \text{ sin } 2\pi\text{ft}$, where I is the current at time t, and $I_0 = V_0/\text{R}$ is the peak current.
- The average AC power is $P_{\text{ave}} = \frac{1}{2} I_0 V_0$.
- Average (rms) current I_{rms} and average (rms) voltage V_{rms} are $I_{\text{rms}} = \frac{I_0}{\sqrt{2}}$ and $V_{\text{rms}} = \frac{V_0}{\sqrt{2}}$, where rms stands for root mean square.
- Thus, $P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}}$.
- Ohm’s law for AC is $I_{\text{rms}} = \frac{V_{\text{rms}}}{R}$.
- Expressions for the average power of an AC circuit are $P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}}$, $P_{\text{ave}} = \frac{V_{\text{rms}}^{2}}{R}$, and $P_{\text{ave}} = I_{\text{rms}}^{2} R$, analogous to the expressions for DC circuits.

Conceptual Questions

Exercise:

Problem:

Give an example of a use of AC power other than in the household. Similarly, give an example of a use of DC power other than that supplied by batteries.

Exercise:

Problem:

Why do voltage, current, and power go through zero 120 times per second for 60-Hz AC electricity?

Exercise:

Problem:

You are riding in a train, gazing into the distance through its window. As close objects streak by, you notice that the nearby fluorescent lights make *dashed* streaks. Explain.

Problem Exercises

Exercise:

Problem:

(a) What is the hot resistance of a 25-W light bulb that runs on 120-V AC? (b) If the bulb's operating temperature is 2700°C, what is its resistance at 2600°C?

Exercise:

Problem:

Certain heavy industrial equipment uses AC power that has a peak voltage of 679 V. What is the rms voltage?

Solution:

480 V

Exercise:

Problem:

A certain circuit breaker trips when the rms current is 15.0 A. What is the corresponding peak current?

Exercise:

Problem:

Military aircraft use 400-Hz AC power, because it is possible to design lighter-weight equipment at this higher frequency. What is the time for one complete cycle of this power?

Solution:

2.50 ms

Exercise:

Problem:

A North American tourist takes his 25.0-W, 120-V AC razor to Europe, finds a special adapter, and plugs it into 240 V AC. Assuming constant resistance, what power does the razor consume as it is ruined?

Exercise:

Problem:

In this problem, you will verify statements made at the end of the power losses for [link]. (a) What current is needed to transmit 100 MW of power at a voltage of 25.0 kV? (b) Find the power loss in a 1.00 - Ω transmission line. (c) What percent loss does this represent?

Solution:

(a) 4.00 kA

(b) 16.0 MW

(c) 16.0%

Exercise:

Problem:

A small office-building air conditioner operates on 408-V AC and consumes 50.0 kW. (a) What is its effective resistance? (b) What is the cost of running the air conditioner during a hot summer month when it is on 8.00 h per day for 30 days and electricity costs $9.00\ \text{cents/kW}\cdot\text{h}$?

Exercise:

Problem:

What is the peak power consumption of a 120-V AC microwave oven that draws 10.0 A?

Solution:

2.40 kW

Exercise:

Problem:

What is the peak current through a 500-W room heater that operates on 120-V AC power?

Exercise:

Problem:

Two different electrical devices have the same power consumption, but one is meant to be operated on 120-V AC and the other on 240-V AC. (a) What is the ratio of their resistances? (b) What is the ratio of their currents? (c) Assuming its resistance is unaffected, by what factor will the power increase if a 120-V AC device is connected to 240-V AC?

Solution:

(a) 4.0

(b) 0.50

(c) 4.0

Exercise:

Problem:

Nichrome wire is used in some radiative heaters. (a) Find the resistance needed if the average power output is to be 1.00 kW utilizing 120-V AC. (b) What length of Nichrome wire, having a cross-sectional area of 5.00mm^2, is needed if the operating temperature is $500^\text{o}\ \text{C}$? (c) What power will it draw when first switched on?

Exercise:

Problem:

Find the time after $t = 0$ when the instantaneous voltage of 60-Hz AC first reaches the following values: (a) $V_0/2$ (b) V_0 (c) 0.

Solution:

(a) 1.39 ms

(b) 4.17 ms

(c) 8.33 ms

Exercise:

Problem:

(a) At what two times in the first period following $t = 0$ does the instantaneous voltage in 60-Hz AC equal V_rms? (b) $-V_\text{rms}$?

Glossary

direct current
: (DC) the flow of electric charge in only one direction

alternating current
: (AC) the flow of electric charge that periodically reverses direction

AC voltage

voltage that fluctuates sinusoidally with time, expressed as $V = V_0 \sin 2\pi f t$, where V is the voltage at time t, V_0 is the peak voltage, and f is the frequency in hertz

AC current

current that fluctuates sinusoidally with time, expressed as $I = I_0 \sin 2\pi f t$, where I is the current at time t, I_0 is the peak current, and f is the frequency in hertz

rms current

the root mean square of the current, $I_{\text{rms}} = I_0/\sqrt{2}$, where I_0 is the peak current, in an AC system

rms voltage

the root mean square of the voltage, $V_{\text{rms}} = V_0/\sqrt{2}$, where V_0 is the peak voltage, in an AC system

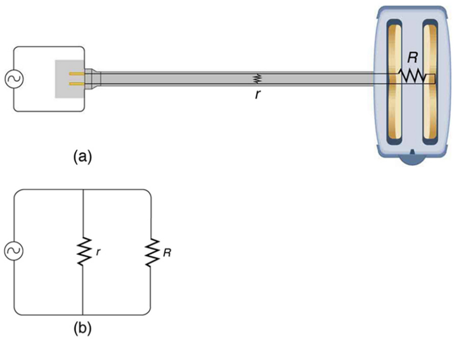
Electric Hazards and the Human Body

- Define thermal hazard, shock hazard, and short circuit.
- Explain what effects various levels of current have on the human body.

There are two known hazards of electricity—thermal and shock. A **thermal hazard** is one where excessive electric power causes undesired thermal effects, such as starting a fire in the wall of a house. A **shock hazard** occurs when electric current passes through a person. Shocks range in severity from painful, but otherwise harmless, to heart-stopping lethality. This section considers these hazards and the various factors affecting them in a quantitative manner. Electrical Safety: Systems and Devices will consider systems and devices for preventing electrical hazards.

Thermal Hazards

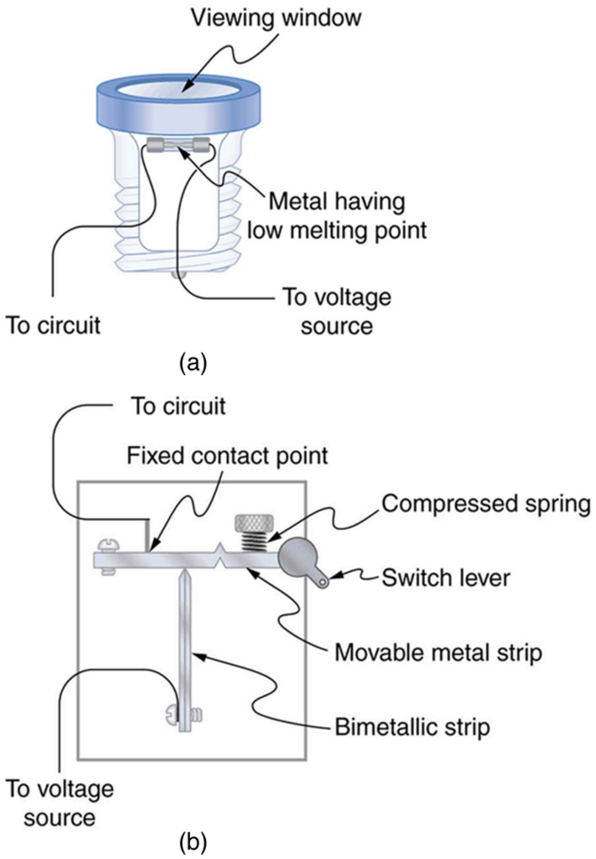
Electric power causes undesired heating effects whenever electric energy is converted to thermal energy at a rate faster than it can be safely dissipated. A classic example of this is the **short circuit**, a low-resistance path between terminals of a voltage source. An example of a short circuit is shown in [link]. Insulation on wires leading to an appliance has worn through, allowing the two wires to come into contact. Such an undesired contact with a high voltage is called a *short*. Since the resistance of the short, r, is very small, the power dissipated in the short, $P=V^2/r$, is very large. For example, if V is 120 V and r is $0.100\ \Omega$, then the power is 144 kW, *much* greater than that used by a typical household appliance. Thermal energy delivered at this rate will very quickly raise the temperature of surrounding materials, melting or perhaps igniting them.



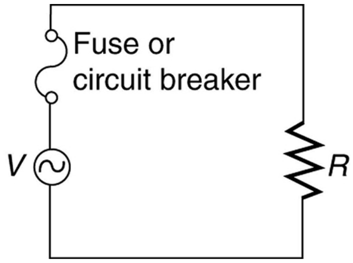
A short circuit is an undesired low-resistance path across a voltage source. (a) Worn insulation on the wires of a toaster allow them to come into contact with a low resistance r. Since $P = V^2/r$, thermal power is created so rapidly that the cord melts or burns. (b) A schematic of the short circuit.

One particularly insidious aspect of a short circuit is that its resistance may actually be decreased due to the increase in temperature. This can happen if the short creates ionization. These charged atoms and molecules are free to move and, thus, lower the resistance r. Since $P = V^2/r$, the power dissipated in the short rises, possibly causing more ionization, more power, and so on. High voltages, such as the 480-V AC used in some industrial applications, lend themselves to this hazard, because higher voltages create higher initial power production in a short.

Another serious, but less dramatic, thermal hazard occurs when wires supplying power to a user are overloaded with too great a current. As discussed in the previous section, the power dissipated in the supply wires is $P = I^2 R_{\mathrm{w}}$, where R_{w} is the resistance of the wires and I the current flowing through them. If either I or R_{w} is too large, the wires overheat. For example, a worn appliance cord (with some of its braided wires broken) may have $R_{\mathrm{w}} = 2.00\ \Omega$ rather than the $0.100\ \Omega$ it should be. If 10.0 A of current passes through the cord, then $P = I^2 R_{\mathrm{w}} = 200\ \mathrm{W}$ is dissipated in the cord—much more than is safe. Similarly, if a wire with a 0.100 - Ω resistance is meant to carry a few amps, but is instead carrying 100 A, it will severely overheat. The power dissipated in the wire will in that case be $P = 1000\ \mathrm{W}$. Fuses and circuit breakers are used to limit excessive currents. (See [link] and [link].) Each device opens the circuit automatically when a sustained current exceeds safe limits.


Viewing window
Metal having low melting point
To circuit
To voltage source
(a)
To circuit
Fixed contact point
Compressed spring
Switch lever
Movable metal strip
Bimetallic strip
To voltage source
(b)

(a) A fuse has a metal strip with a low melting point that, when overheated by an excessive current, permanently breaks the connection of a circuit to a voltage source. (b) A circuit breaker is an automatic but restorable electric switch. The one shown here has a bimetallic strip that bends to the right and into the notch if overheated. The spring then forces the metal strip downward, breaking the electrical connection at the points.



Schematic of a circuit with a fuse or circuit breaker in it. Fuses and circuit breakers act like automatic switches that open when sustained current exceeds desired limits.

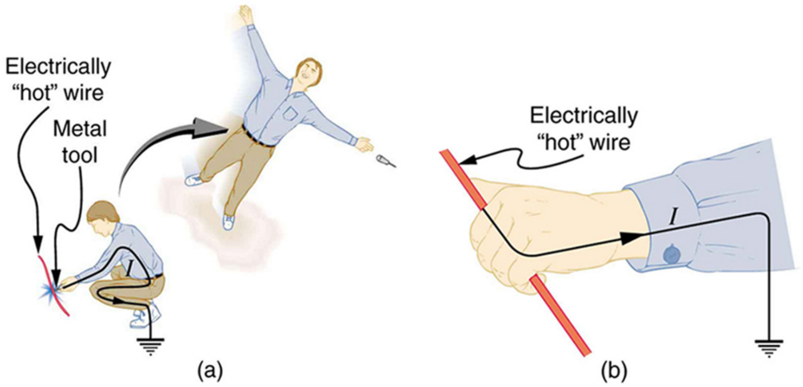
Fuses and circuit breakers for typical household voltages and currents are relatively simple to produce, but those for large voltages and currents experience special problems. For example, when a circuit breaker tries to interrupt the flow of high-voltage electricity, a spark can jump across its points that ionizes the air in the gap and allows the current to continue flowing. Large circuit breakers found in power-distribution systems employ insulating gas and even use jets of gas to blow out such sparks. Here AC is safer than DC, since AC current goes through zero 120 times per second, giving a quick opportunity to extinguish these arcs.

Shock Hazards

Electrical currents through people produce tremendously varied effects. An electrical current can be used to block back pain. The possibility of using electrical current to stimulate muscle action in paralyzed limbs, perhaps allowing paraplegics to walk, is under study. TV dramatizations in which electrical shocks are used to bring a heart attack victim out of ventricular fibrillation (a massively irregular, often fatal, beating of the heart) are more than common. Yet most electrical shock fatalities occur because a current put the heart into fibrillation. A pacemaker uses electrical shocks to stimulate the heart to beat properly. Some fatal shocks do not produce burns, but warts can be safely burned off with electric current (though freezing using liquid nitrogen is now more common). Of course, there are consistent explanations for these disparate effects. The major factors upon which the effects of electrical shock depend are

1. The amount of current I
2. The path taken by the current
3. The duration of the shock
4. The frequency f of the current ($f = 0$ for DC)

[link] gives the effects of electrical shocks as a function of current for a typical accidental shock. The effects are for a shock that passes through the trunk of the body, has a duration of 1 s, and is caused by 60-Hz power.


Electrically
"hot" wire
Metal
tool
I
(a)
Electrically
"hot" wire
I
(b)

An electric current can cause muscular contractions with varying effects. (a) The victim is "thrown" backward by involuntary muscle contractions that extend the legs and torso. (b) The victim can't let go of the wire that is stimulating all the muscles in the hand. Those that close the fingers are stronger than those that open them.

Current (mA)	**Effect**
1	Threshold of sensation
5	Maximum harmless current
10–20	Onset of sustained muscular contraction; cannot let go for duration of shock; contraction of chest muscles may stop breathing during shock

Current (mA)	**Effect**
50	Onset of pain
100–300+	Ventricular fibrillation possible; often fatal
300	Onset of burns depending on concentration of current
6000 (6 A)	Onset of sustained ventricular contraction and respiratory paralysis; both cease when shock ends; heartbeat may return to normal; used to defibrillate the heart

Effects of Electrical Shock as a Function of Current[footnote]
For an average male shocked through trunk of body for 1 s by 60-Hz AC. Values for females are 60–80% of those listed.

Our bodies are relatively good conductors due to the water in our bodies. Given that larger currents will flow through sections with lower resistance (to be further discussed in the next chapter), electric currents preferentially flow through paths in the human body that have a minimum resistance in a direct path to earth. The earth is a natural electron sink. Wearing insulating shoes, a requirement in many professions, prohibits a pathway for electrons by providing a large resistance in that path. Whenever working with high-power tools (drills), or in risky situations, ensure that you do not provide a pathway for current flow (especially through the heart).

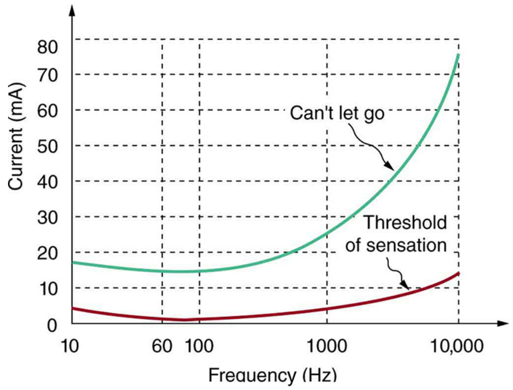
Very small currents pass harmlessly and unfelt through the body. This happens to you regularly without your knowledge. The threshold of sensation is only 1 mA and, although unpleasant, shocks are apparently harmless for currents less than 5 mA. A great number of safety rules take the 5-mA value for the maximum allowed shock. At 10 to 20 mA and above, the current can stimulate sustained muscular contractions much as regular nerve impulses do. People sometimes say they were knocked across the room by a shock, but what really happened was that certain muscles contracted, propelling them in a manner not of their own choosing. (See [link](a).) More frightening, and potentially more dangerous, is the “can’t let go” effect illustrated in [link](b). The muscles that close the fingers are stronger than those that open them, so the hand closes involuntarily on the wire shocking it. This can prolong the shock indefinitely. It can also be a danger to a person trying to rescue the victim, because the rescuer’s hand may close about the victim’s wrist. Usually the best way to help the victim is to give the fist a hard knock/blow/jar with an insulator or to throw an insulator at the fist. Modern electric fences, used in animal enclosures, are now pulsed on and off to allow people who touch them to get free, rendering them less lethal than in the past.

Greater currents may affect the heart. Its electrical patterns can be disrupted, so that it beats irregularly and ineffectively in a condition called “ventricular fibrillation.” This condition often lingers after the shock and is fatal due to a lack of blood circulation. The threshold for ventricular fibrillation is between 100 and 300 mA. At about 300 mA and above, the shock can cause burns, depending on the concentration of current—the more concentrated, the greater the likelihood of burns.

Very large currents cause the heart and diaphragm to contract for the duration of the shock. Both the heart and breathing stop. Interestingly, both often return to normal following the shock. The electrical patterns on the heart are completely erased in a manner that the heart can start afresh with normal beating, as opposed to the permanent disruption caused by smaller currents that can put the heart into ventricular fibrillation. The latter is something like scribbling on a blackboard, whereas the former completely erases it. TV dramatizations of electric shock used to bring a heart attack victim out of ventricular fibrillation also show large paddles. These are used to spread out current passed through the victim to reduce the likelihood of burns.

Current is the major factor determining shock severity (given that other conditions such as path, duration, and frequency are fixed, such as in the table and preceding discussion). A larger voltage is more hazardous, but since $I = \text{V}/\text{R}$, the severity of the shock depends on the combination of voltage and resistance. For example, a person with dry skin has a resistance of about $200\ \text{k}\Omega$. If he comes into contact with 120-V AC, a current $I=(120\ \text{V})/(200\ \text{k}\Omega)=0.6\ \text{mA}$ passes harmlessly through him. The same person soaking wet may have a resistance of $10.0\ \text{k}\Omega$ and the same 120 V will produce a current of 12 mA—above the “can’t let go” threshold and potentially dangerous.

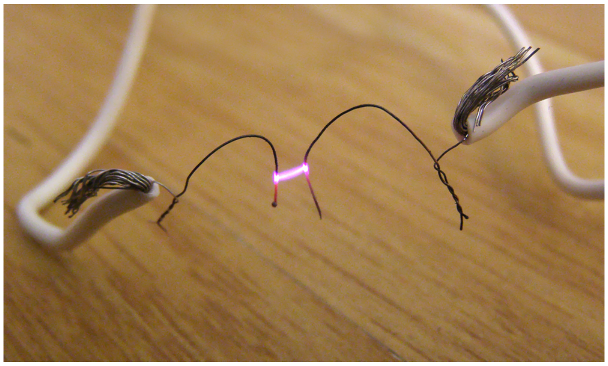
Most of the body’s resistance is in its dry skin. When wet, salts go into ion form, lowering the resistance significantly. The interior of the body has a much lower resistance than dry skin because of all the ionic solutions and fluids it contains. If skin resistance is bypassed, such as by an intravenous infusion, a catheter, or exposed pacemaker leads, a person is rendered **microshock sensitive**. In this condition, currents about 1/1000 those listed in [link] produce similar effects. During open-heart surgery, currents as small as $20\ \mu\text{A}$ can be used to still the heart. Stringent electrical safety requirements in hospitals, particularly in surgery and intensive care, are related to the doubly disadvantaged microshock-sensitive patient. The break in the skin has reduced his resistance, and so the same voltage causes a greater current, and a much smaller current has a greater effect.



Graph of average values for the threshold of sensation and the “can’t let go” current as a function of frequency. The lower the value, the

more sensitive the body is
at that frequency.

Factors other than current that affect the severity of a shock are its path, duration, and AC frequency. Path has obvious consequences. For example, the heart is unaffected by an electric shock through the brain, such as may be used to treat manic depression. And it is a general truth that the longer the duration of a shock, the greater its effects. [link] presents a graph that illustrates the effects of frequency on a shock. The curves show the minimum current for two different effects, as a function of frequency. The lower the current needed, the more sensitive the body is at that frequency. Ironically, the body is most sensitive to frequencies near the 50- or 60-Hz frequencies in common use. The body is slightly less sensitive for DC ($f=0$), mildly confirming Edison's claims that AC presents a greater hazard. At higher and higher frequencies, the body becomes progressively less sensitive to any effects that involve nerves. This is related to the maximum rates at which nerves can fire or be stimulated. At very high frequencies, electrical current travels only on the surface of a person. Thus a wart can be burned off with very high frequency current without causing the heart to stop. (Do not try this at home with 60-Hz AC!) Some of the spectacular demonstrations of electricity, in which high-voltage arcs are passed through the air and over people's bodies, employ high frequencies and low currents. (See [link].) Electrical safety devices and techniques are discussed in detail in Electrical Safety: Systems and Devices.



Is this electric arc dangerous?

The answer depends on the AC frequency and the power involved. (credit: Khimich Alex, Wikimedia Commons)

Section Summary

- The two types of electric hazards are thermal (excessive power) and shock (current through a person).
- Shock severity is determined by current, path, duration, and AC frequency.
- [link] lists shock hazards as a function of current.
- [link] graphs the threshold current for two hazards as a function of frequency.

Conceptual Questions

Exercise:

Problem:

Using an ohmmeter, a student measures the resistance between various points on his body. He finds that the resistance between two points on the same finger is about the same as the resistance between two points on opposite hands—both are several hundred thousand ohms. Furthermore, the resistance decreases when more skin is brought into contact with the probes of the ohmmeter. Finally, there is a dramatic drop in resistance (to a few thousand ohms) when the skin is wet. Explain these observations and their implications regarding skin and internal resistance of the human body.

Exercise:

Problem: What are the two major hazards of electricity?

Exercise:

Problem: Why isn’t a short circuit a shock hazard?

Exercise:

Problem:

What determines the severity of a shock? Can you say that a certain voltage is hazardous without further information?

Exercise:

Problem:

An electrified needle is used to burn off warts, with the circuit being completed by having the patient sit on a large butt plate. Why is this plate large?

Exercise:

Problem:

Some surgery is performed with high-voltage electricity passing from a metal scalpel through the tissue being cut. Considering the nature of electric fields at the surface of conductors, why would you expect most of the current to flow from the sharp edge of the scalpel? Do you think high- or low-frequency AC is used?

Exercise:

Problem:

Some devices often used in bathrooms, such as hairdryers, often have safety messages saying “Do not use when the bathtub or basin is full of water.” Why is this so?

Exercise:

Problem:

We are often advised to not flick electric switches with wet hands, dry your hand first. We are also advised to never throw water on an electric fire. Why is this so?

Exercise:

Problem:

Before working on a power transmission line, linemen will touch the line with the back of the hand as a final check that the voltage is zero. Why the back of the hand?

Exercise:

Problem:

Why is the resistance of wet skin so much smaller than dry, and why do blood and other bodily fluids have low resistances?

Exercise:

Problem:

Could a person on intravenous infusion (an IV) be microshock sensitive?

Exercise:

Problem:

In view of the small currents that cause shock hazards and the larger currents that circuit breakers and fuses interrupt, how do they play a role in preventing shock hazards?

Problem Exercises

Exercise:

Problem:

(a) How much power is dissipated in a short circuit of 240-V AC through a resistance of $0.250\ \Omega$? (b) What current flows?

Solution:

(a) 230 kW

(b) 960 A

Exercise:

Problem:

What voltage is involved in a 1.44-kW short circuit through a $0.100 - \Omega$ resistance?

Exercise:

Problem:

Find the current through a person and identify the likely effect on her if she touches a 120-V AC source: (a) if she is standing on a rubber mat and offers a total resistance of $300\text{ k}\Omega$; (b) if she is standing barefoot on wet grass and has a resistance of only $4000\text{ k}\Omega$.

Solution:

(a) 0.400 mA, no effect

(b) 26.7 mA, muscular contraction for duration of the shock (can't let go)

Exercise:

Problem:

While taking a bath, a person touches the metal case of a radio. The path through the person to the drainpipe and ground has a resistance of $4000\ \Omega$. What is the smallest voltage on the case of the radio that could cause ventricular fibrillation?

Exercise:

Problem:

Foolishly trying to fish a burning piece of bread from a toaster with a metal butter knife, a man comes into contact with 120-V AC. He does not even feel it since, luckily, he is wearing rubber-soled shoes. What is the minimum resistance of the path the current follows through the person?

Solution:

$1.20\times10^{5}\ \Omega$

Exercise:

Problem:

(a) During surgery, a current as small as $20.0\ \mu\text{A}$ applied directly to the heart may cause ventricular fibrillation. If the resistance of the exposed heart is $300\ \Omega$, what is the smallest voltage that poses this danger? (b) Does your answer imply that special electrical safety precautions are needed?

Exercise:

Problem:

(a) What is the resistance of a 220-V AC short circuit that generates a peak power of 96.8 kW? (b) What would the average power be if the voltage was 120 V AC?

Solution:

(a) $1.00\ \Omega$

(b) 14.4 kW

Exercise:

Problem:

A heart defibrillator passes 10.0 A through a patient’s torso for 5.00 ms in an attempt to restore normal beating. (a) How much charge passed? (b) What voltage was applied if 500 J of energy was dissipated? (c) What was the path’s resistance? (d) Find the temperature increase caused in the 8.00 kg of affected tissue.

Exercise:

Problem: **Integrated Concepts**

A short circuit in a 120-V appliance cord has a $0.500\text{-}\Omega$ resistance. Calculate the temperature rise of the 2.00 g of surrounding materials, assuming their specific heat capacity is $0.200\ \text{cal/g}\cdot^{\text{o}}\ \text{C}$ and that it takes 0.0500 s for a circuit breaker to interrupt the current. Is this likely to be damaging?

Solution:

Temperature increases $860^{\text{o}}\ \text{C}$. It is very likely to be damaging.

Exercise:

Problem: **Construct Your Own Problem**

Consider a person working in an environment where electric currents might pass through her body. Construct a problem in which you calculate the resistance of insulation needed to protect the person from harm. Among the things to be considered are the voltage to which the person might be exposed, likely body resistance (dry, wet, …), and acceptable currents (safe but sensed, safe and unfelt, …).

Glossary

thermal hazard	a hazard in which electric current causes undesired thermal effects

shock hazard
when electric current passes through a person

short circuit
also known as a “short,” a low-resistance path between terminals of a voltage source

microshock sensitive
a condition in which a person’s skin resistance is bypassed, possibly by a medical procedure, rendering the person vulnerable to electrical shock at currents about 1/1000 the normally required level

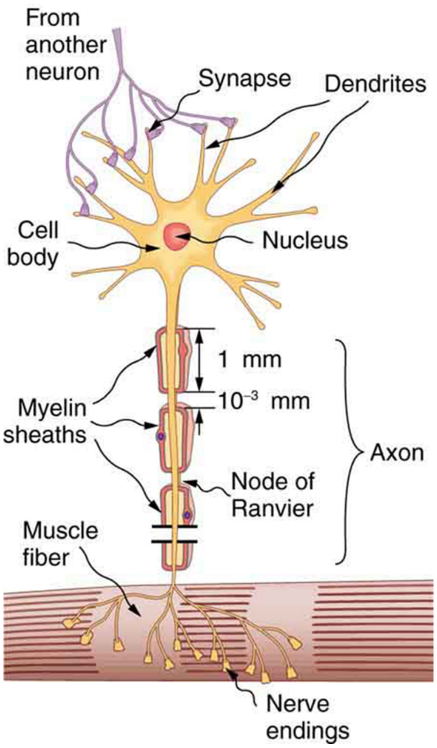
Nerve Conduction–Electrocardiograms

- Explain the process by which electric signals are transmitted along a neuron.
- Explain the effects myelin sheaths have on signal propagation.
- Explain what the features of an ECG signal indicate.

Nerve Conduction

Electric currents in the vastly complex system of billions of nerves in our body allow us to sense the world, control parts of our body, and think. These are representative of the three major functions of nerves. First, nerves carry messages from our sensory organs and others to the central nervous system, consisting of the brain and spinal cord. Second, nerves carry messages from the central nervous system to muscles and other organs. Third, nerves transmit and process signals within the central nervous system. The sheer number of nerve cells and the incredibly greater number of connections between them makes this system the subtle wonder that it is. **Nerve conduction** is a general term for electrical signals carried by nerve cells. It is one aspect of **bioelectricity**, or electrical effects in and created by biological systems.

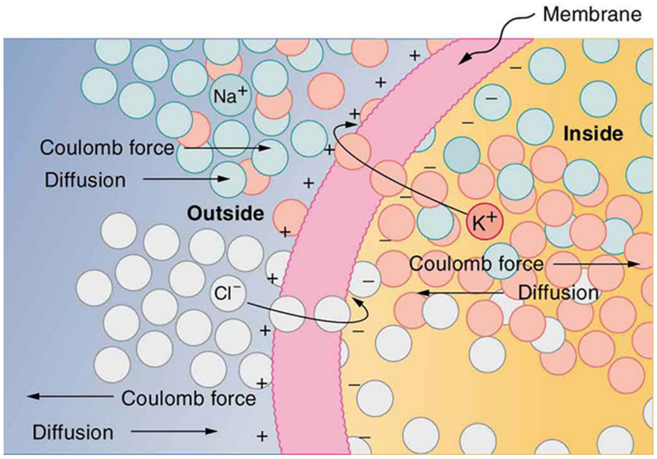
Nerve cells, properly called *neurons*, look different from other cells—they have tendrils, some of them many centimeters long, connecting them with other cells. (See [link].) Signals arrive at the cell body across *synapses* or through *dendrites*, stimulating the neuron to generate its own signal, sent along its long *axon* to other nerve or muscle cells. Signals may arrive from many other locations and be transmitted to yet others, conditioning the synapses by use, giving the system its complexity and its ability to learn.


From another neuron
Synapse
Dendrites
Cell body
Nucleus
1 mm
10^{-3} mm
Myelin sheaths
Axon
Node of Ranvier
Muscle fiber
Nerve endings

A neuron with its dendrites and long axon. Signals in the form of electric currents reach the cell body through dendrites and across synapses, stimulating the neuron to generate its own signal sent down the axon. The number of interconnections can be far greater than shown here.

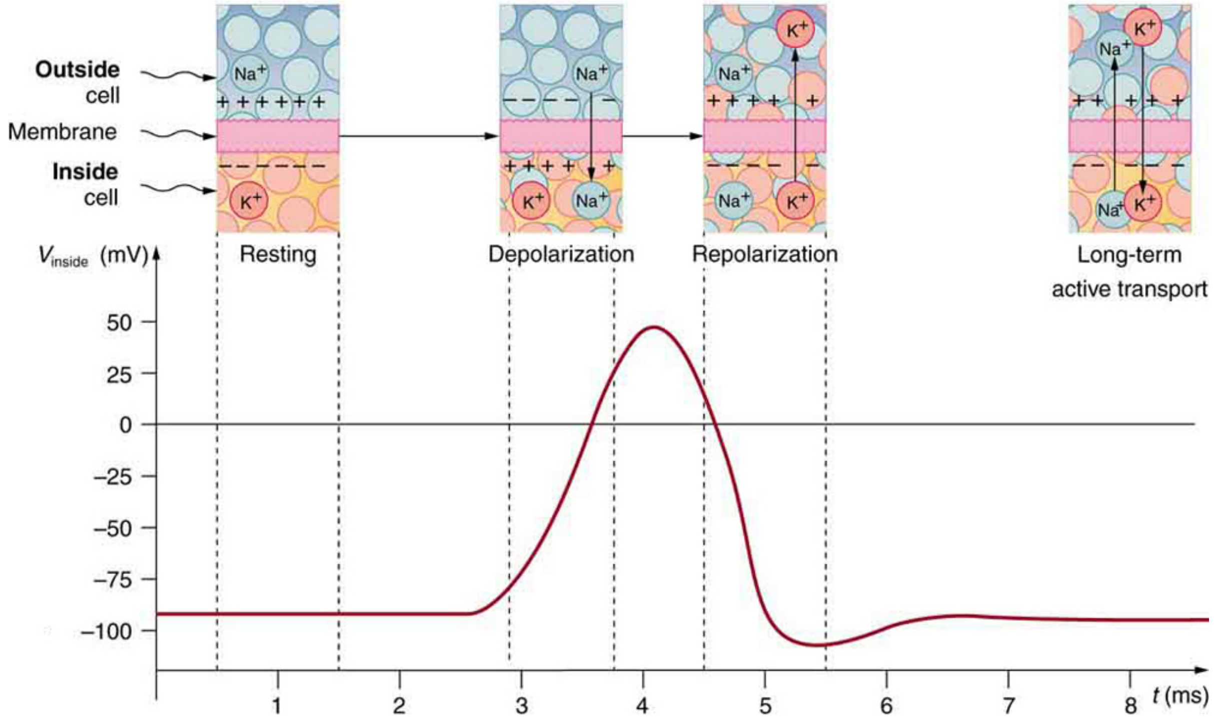
The method by which these electric currents are generated and transmitted is more complex than the simple movement of free charges in a conductor, but it can be understood with principles already discussed in this text. The most important of these are the Coulomb force and diffusion.

[link] illustrates how a voltage (potential difference) is created across the cell membrane of a neuron in its resting state. This thin membrane separates electrically neutral fluids having differing concentrations of ions, the most important varieties being Na^+, K^+, and Cl^- (these are sodium, potassium, and chlorine ions with single plus or minus charges as indicated). As discussed in Molecular Transport Phenomena: Diffusion, Osmosis, and Related Processes, free ions will diffuse from a region of high concentration to one of low concentration. But the cell membrane is **semipermeable**, meaning that some ions may cross it while others cannot. In its resting state, the cell membrane is permeable to K^+ and Cl^-, and impermeable to Na^+. Diffusion of K^+ and Cl^- thus creates the layers of positive and negative charge on the outside and inside of the membrane. The Coulomb force prevents the ions from diffusing across in their entirety. Once the charge layer has built up, the repulsion of like charges prevents more from moving across, and the attraction of unlike charges prevents more from leaving either side. The result is two layers of charge right on the membrane, with diffusion being balanced by the Coulomb force. A tiny fraction of the charges move across and the fluids remain neutral (other ions are present), while a separation of charge and a voltage have been created across the membrane.



The semipermeable membrane of a

cell has different concentrations of ions inside and out. Diffusion moves the K^+ and Cl^- ions in the direction shown, until the Coulomb force halts further transfer. This results in a layer of positive charge on the outside, a layer of negative charge on the inside, and thus a voltage across the cell membrane. The membrane is normally impermeable to Na^+.

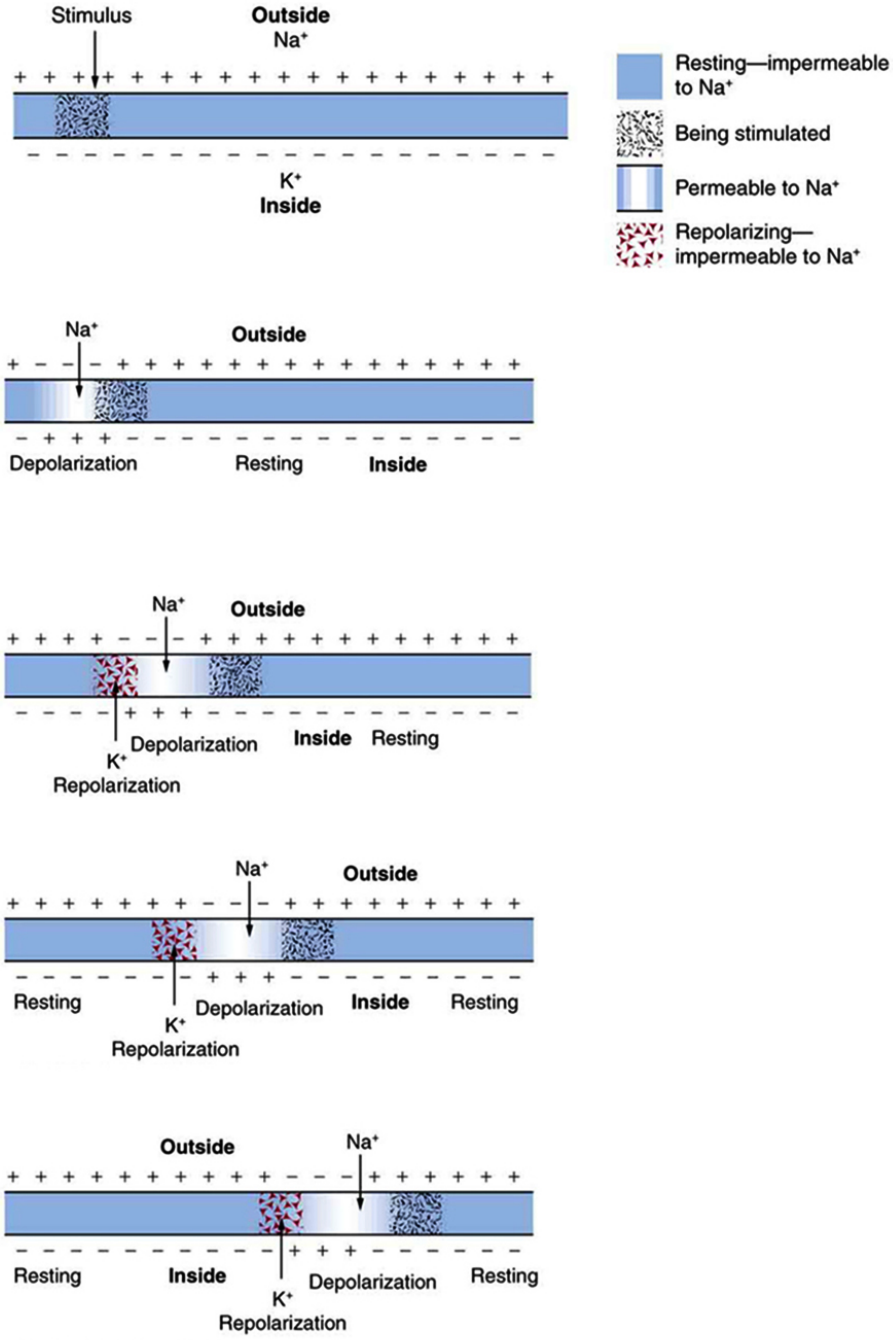


An action potential is the pulse of voltage inside a nerve cell graphed here. It is caused by movements of ions across the cell membrane as shown. Depolarization occurs when a stimulus makes the membrane permeable to Na^+ ions. Repolarization follows as the membrane again becomes impermeable to Na^+, and K^+ moves from high to low concentration. In the long term, active transport slowly maintains the concentration differences, but the cell may fire hundreds of times in rapid succession without seriously depleting them.

The separation of charge creates a potential difference of 70 to 90 mV across the cell membrane. While this is a small voltage, the resulting electric field ($E = V/d$) across the only 8-nm-thick membrane is immense (on the order of 11 MV/m!) and has fundamental effects on its structure and permeability. Now, if the exterior of a neuron is taken to be at 0 V, then the interior has a *resting potential* of about –90 mV. Such voltages are created across the membranes of almost all types of animal cells but are largest in nerve and muscle cells. In fact, fully 25% of the energy used by cells goes toward creating and maintaining these potentials.

Electric currents along the cell membrane are created by any stimulus that changes the membrane’s permeability. The membrane thus temporarily becomes permeable to Na^+, which then rushes in, driven both by diffusion and the Coulomb force. This inrush of Na^+ first neutralizes the inside membrane, or *depolarizes* it, and then makes it slightly positive. The depolarization causes the membrane to again become impermeable to Na^+, and the movement of K^+ quickly returns the cell to its resting potential, or *repolarizes* it. This sequence of events results in a voltage pulse, called the *action potential*. (See [link].) Only small fractions of the ions move, so that the cell can fire many hundreds of times without depleting the excess concentrations of Na^+ and K^+. Eventually, the cell must replenish these ions to maintain the concentration differences that create bioelectricity. This sodium-potassium pump is an example of *active transport*, wherein cell energy is used to move ions across membranes against diffusion gradients and the Coulomb force.

The action potential is a voltage pulse at one location on a cell membrane. How does it get transmitted along the cell membrane, and in particular down an axon, as a nerve impulse? The answer is that the changing voltage and electric fields affect the permeability of the adjacent cell membrane, so that the same process takes place there. The adjacent membrane depolarizes, affecting the membrane further down, and so on, as illustrated in [link]. Thus the action potential stimulated at one location triggers a *nerve impulse* that moves slowly (about 1 m/s) along the cell membrane.



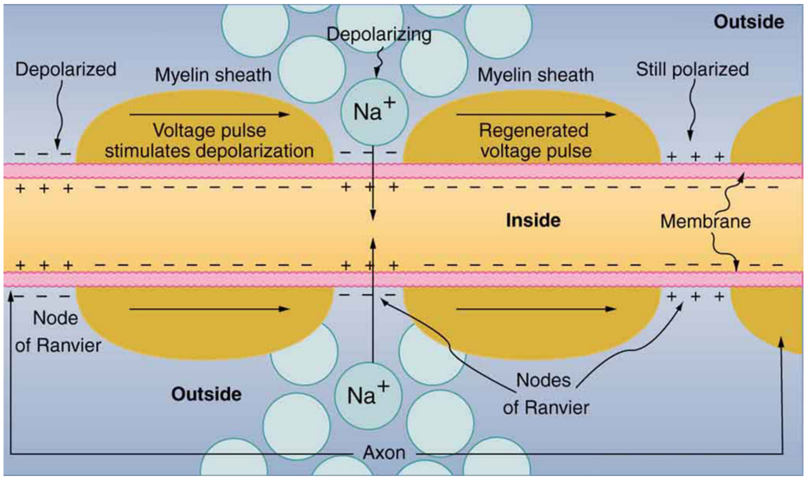
Stimulus
Outside
Na^+
K^+
Inside
Resting—impermeable to Na^+
Being stimulated
Permeable to Na^+
Repolarizing—impermeable to Na^+
Na^+
Outside
Depolarization
Resting
Inside
Na^+
Outside
Depolarization
Inside
Resting
K^+
Repolarization
Na^+
Outside
Resting
Depolarization
Inside
Resting
K^+
Repolarization
Outside
Na^+
Resting
Inside
Depolarization
Resting
K^+
Repolarization

A nerve impulse is the propagation of an action potential along a cell membrane. A stimulus causes an action potential at one location, which changes the permeability of the adjacent membrane, causing an action potential there. This in turn affects the membrane further down, so that the action potential moves slowly (in electrical terms) along the cell membrane. Although the impulse is due to Na^+ and K^+ going across the membrane, it is equivalent to a wave of charge moving along the outside and inside of the membrane.

Some axons, like that in [link], are sheathed with *myelin*, consisting of fat-containing cells. [link] shows an enlarged view of an axon having myelin sheaths characteristically separated by unmyelinated gaps (called nodes of Ranvier). This arrangement gives the axon a number of interesting properties. Since myelin is an insulator, it prevents signals from jumping between adjacent nerves (cross talk). Additionally, the myelinated regions transmit electrical signals at a very high speed, as an ordinary conductor or resistor would. There is no action potential in the myelinated regions, so that no cell energy is used in them. There is an IR signal loss in the myelin, but the signal is regenerated in the gaps, where the voltage pulse triggers the action potential at full voltage. So a myelinated axon transmits a nerve impulse faster, with less energy consumption, and is better protected from cross talk than an unmyelinated one. Not all axons are myelinated, so that cross talk and slow signal transmission are a characteristic of the normal operation of these axons, another variable in the nervous system.

The degeneration or destruction of the myelin sheaths that surround the nerve fibers impairs signal transmission and can lead to numerous neurological effects. One of the most prominent of these diseases comes from the body’s own immune system attacking the myelin in the central nervous system—multiple sclerosis. MS symptoms include fatigue, vision problems, weakness of arms and legs, loss of balance, and tingling or numbness in one’s extremities (neuropathy). It is more apt to strike younger adults, especially females. Causes might come from infection, environmental or geographic affects, or genetics. At the moment there is no known cure for MS.

Most animal cells can fire or create their own action potential. Muscle cells contract when they fire and are often induced to do so by a nerve impulse. In fact, nerve and muscle cells are physiologically similar, and there are even hybrid cells, such as in the heart, that have characteristics of both nerves and muscles. Some animals, like the infamous electric eel (see [link]), use muscles ganged so that their voltages add in order to create a shock great enough to stun prey.

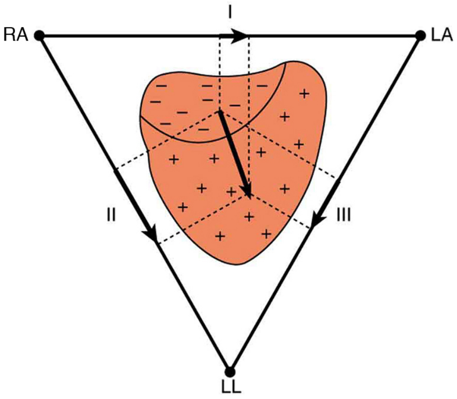

Outside
Depolarizing
Depolarized
Myelin sheath
Myelin sheath
Still polarized
Na+
Voltage pulse stimulates depolarization
Regenerated voltage pulse
Inside
Membrane
Node of Ranvier
Outside
Na+
Nodes of Ranvier
Axon

Propagation of a nerve impulse down a myelinated axon, from left to right. The signal travels very fast and without energy input in the myelinated regions, but it loses voltage. It is regenerated in the gaps. The signal moves faster than in unmyelinated axons and is insulated from signals in other nerves, limiting cross talk.

An electric eel flexes its muscles to create a voltage that stuns prey. (credit: chrisbb, Flickr)

Electrocardiograms

Just as nerve impulses are transmitted by depolarization and repolarization of adjacent membrane, the depolarization that causes muscle contraction can also stimulate adjacent muscle cells to depolarize (fire) and contract. Thus, a depolarization wave can be sent across the heart, coordinating its rhythmic contractions and enabling it to perform its vital function of propelling blood through the circulatory system. [link] is a simplified graphic of a depolarization wave spreading across the heart from the *sinoarterial (SA) node*, the heart’s natural pacemaker.

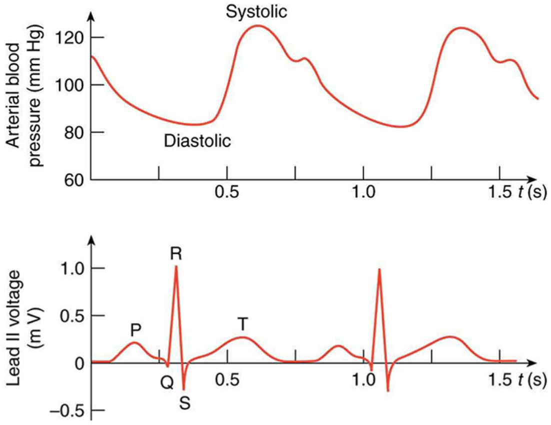


The outer surface of the heart changes from positive to negative during depolarization. This wave of depolarization is spreading from the top of the heart and is represented by a vector pointing in the direction of the wave. This vector is a voltage (potential difference) vector. Three electrodes, labeled RA, LA, and LL, are placed on the patient. Each pair (called leads I, II, and III) measures a component of the depolarization vector and is graphed in an ECG.

An **electrocardiogram (ECG)** is a record of the voltages created by the wave of depolarization and subsequent repolarization in the heart. Voltages between pairs of electrodes placed on the chest are vector components of the voltage wave on the heart. Standard ECGs have 12 or more electrodes, but only three are shown in [link] for clarity. Decades ago, three-electrode ECGs were performed by placing electrodes on the left and right arms and the left leg. The voltage between the right arm and the left leg is called the *lead II potential* and is the most often graphed. We shall examine the lead II potential as an indicator of heart-muscle function and see that it is coordinated with arterial blood pressure as well.

Heart function and its four-chamber action are explored in Viscosity and Laminar Flow; Poiseuille's Law. Basically, the right and left atria receive blood from the body and lungs, respectively, and pump the blood into the ventricles. The right and left ventricles, in turn, pump blood through the lungs and the rest of the body, respectively. Depolarization of the heart muscle causes it to contract. After contraction it is repolarized to ready it for the next beat. The ECG measures components of depolarization and repolarization of the heart muscle and can yield significant information on the functioning and malfunctioning of the heart.

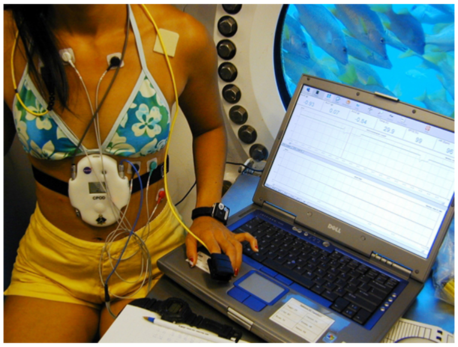
[link] shows an ECG of the lead II potential and a graph of the corresponding arterial blood pressure. The major features are labeled P, Q, R, S, and T. The *P wave* is generated by the depolarization and contraction of the atria as they pump blood into the ventricles. The *QRS complex* is created by the depolarization of the ventricles as they pump blood to the lungs and body. Since the shape of the heart and the path of the depolarization wave are not simple, the QRS complex has this typical shape and time span. The lead II QRS signal also masks the repolarization of the atria, which occur at the same time. Finally, the *T wave* is generated by the repolarization of the ventricles and is followed by the next P wave in the next heartbeat. Arterial blood pressure varies with each part of the heartbeat, with systolic (maximum) pressure occurring closely after the QRS complex, which signals contraction of the ventricles.



A lead II ECG with

corresponding arterial blood pressure. The QRS complex is created by the depolarization and contraction of the ventricles and is followed shortly by the maximum or systolic blood pressure. See text for further description.

Taken together, the 12 leads of a state-of-the-art ECG can yield a wealth of information about the heart. For example, regions of damaged heart tissue, called infarcts, reflect electrical waves and are apparent in one or more lead potentials. Subtle changes due to slight or gradual damage to the heart are most readily detected by comparing a recent ECG to an older one. This is particularly the case since individual heart shape, size, and orientation can cause variations in ECGs from one individual to another. ECG technology has advanced to the point where a portable ECG monitor with a liquid crystal instant display and a printer can be carried to patients' homes or used in emergency vehicles. See [link].



This NASA scientist and NEEMO 5 aquanaut’s heart rate and other vital signs are being recorded by a portable device while living in an underwater habitat. (credit: NASA, Life Sciences Data Archive at Johnson Space Center, Houston, Texas)

Note:
PhET Explorations: Neuron
Stimulate a neuron and monitor what happens. Pause, rewind, and move forward in time in order to observe the ions as they move across the neuron membrane.
https://phet.colorado.edu/sims/html/neuron/latest/neuron_en.html

Section Summary

- Electric potentials in neurons and other cells are created by ionic concentration differences across semipermeable membranes.
- Stimuli change the permeability and create action potentials that propagate along neurons.
- Myelin sheaths speed this process and reduce the needed energy input.
- This process in the heart can be measured with an electrocardiogram (ECG).

Conceptual Questions

Exercise:

Problem:

Note that in [link], both the concentration gradient and the Coulomb force tend to move Na^+ ions into the cell. What prevents this?

Exercise:

Problem:

Define depolarization, repolarization, and the action potential.

Exercise:

Problem:

Explain the properties of myelinated nerves in terms of the insulating properties of myelin.

Problems & Exercises

Exercise:

Problem: Integrated Concepts

Use the ECG in [link] to determine the heart rate in beats per minute assuming a constant time between beats.

Solution:

80 beats/minute

Exercise:

Problem: Integrated Concepts

(a) Referring to [link], find the time systolic pressure lags behind the middle of the QRS complex. (b) Discuss the reasons for the time lag.

Glossary

nerve conduction
the transport of electrical signals by nerve cells

bioelectricity
electrical effects in and created by biological systems

semipermeable
property of a membrane that allows only certain types of ions to cross it

electrocardiogram (ECG)
usually abbreviated ECG, a record of voltages created by depolarization and repolarization, especially in the heart

Introduction to Circuits and DC Instruments class="introduction"

Electric circuits in a computer allow large amounts of data to be quickly and accurately analyzed.. (credit: Airman 1st Class Mike Meares, United States Air Force)

163:21:51:33
MISSI
BRIE
2 SOPS
MISSION
BRIEF
CENTURIONS
NAVSTAR GPS OPERATIONS

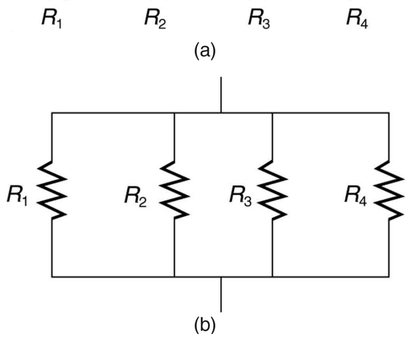
Electric circuits are commonplace. Some are simple, such as those in flashlights. Others, such as those used in supercomputers, are extremely complex.

This collection of modules takes the topic of electric circuits a step beyond simple circuits. When the circuit is purely resistive, everything in this module applies to both DC and AC. Matters become more complex when capacitance is involved. We do consider what happens when capacitors are connected to DC voltage sources, but the interaction of capacitors and other nonresistive devices with AC is left for a later chapter. Finally, a number of important DC instruments, such as meters that measure voltage and current, are covered in this chapter.

Resistors in Series and Parallel

- Draw a circuit with resistors in parallel and in series.
- Calculate the voltage drop of a current across a resistor using Ohm’s law.
- Contrast the way total resistance is calculated for resistors in series and in parallel.
- Explain why total resistance of a parallel circuit is less than the smallest resistance of any of the resistors in that circuit.
- Calculate total resistance of a circuit that contains a mixture of resistors connected in series and in parallel.

Most circuits have more than one component, called a **resistor** that limits the flow of charge in the circuit. A measure of this limit on charge flow is called **resistance**. The simplest combinations of resistors are the series and parallel connections illustrated in [link]. The total resistance of a combination of resistors depends on both their individual values and how they are connected.

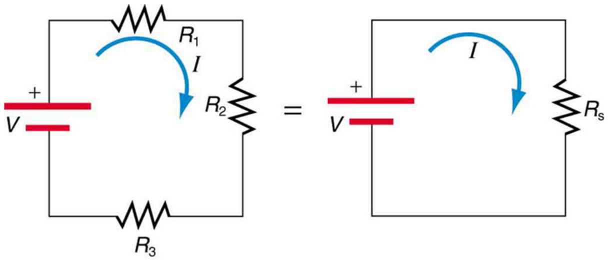


(a) A series connection of resistors. (b) A parallel connection of resistors.

Resistors in Series

When are resistors in **series**? Resistors are in series whenever the flow of charge, called the **current**, must flow through devices sequentially. For example, if current flows through a person holding a screwdriver and into the Earth, then R_1 in [link](a) could be the resistance of the screwdriver's shaft, R_2 the resistance of its handle, R_3 the person's body resistance, and R_4 the resistance of her shoes.

[link] shows resistors in series connected to a **voltage** source. It seems reasonable that the total resistance is the sum of the individual resistances, considering that the current has to pass through each resistor in sequence. (This fact would be an advantage to a person wishing to avoid an electrical shock, who could reduce the current by wearing high-resistance rubber-soled shoes. It could be a disadvantage if one of the resistances were a faulty high-resistance cord to an appliance that would reduce the operating current.)



Three resistors connected in series to a battery (left) and the equivalent single or series resistance (right).

To verify that resistances in series do indeed add, let us consider the loss of electrical power, called a **voltage drop**, in each resistor in [link].

According to **Ohm's law**, the voltage drop, V, across a resistor when a current flows through it is calculated using the equation $V = \mathrm{IR}$, where I equals the current in amps (A) and R is the resistance in ohms (Ω). Another way to think of this is that V is the voltage necessary to make a current I flow through a resistance R.

So the voltage drop across R_1 is $V_1 = IR_1$, that across R_2 is $V_2 = IR_2$, and that across R_3 is $V_3 = IR_3$. The sum of these voltages equals the voltage output of the source; that is,

Equation:

$$V = V_1 + V_2 + V_3.$$

This equation is based on the conservation of energy and conservation of charge. Electrical potential energy can be described by the equation $\text{PE} = \text{qV}$, where q is the electric charge and V is the voltage. Thus the energy supplied by the source is qV, while that dissipated by the resistors is

Equation:

$$qV_1 + qV_2 + qV_3.$$

Note:

Connections: Conservation Laws

The derivations of the expressions for series and parallel resistance are based on the laws of conservation of energy and conservation of charge, which state that total charge and total energy are constant in any process. These two laws are directly involved in all electrical phenomena and will be invoked repeatedly to explain both specific effects and the general behavior of electricity.

These energies must be equal, because there is no other source and no other destination for energy in the circuit. Thus, $qV = qV_1 + qV_2 + qV_3$. The charge q cancels, yielding $V = V_1 + V_2 + V_3$, as stated. (Note that the same amount of charge passes through the battery and each resistor in a given amount of time, since there is no capacitance to store charge, there is no place for charge to leak, and charge is conserved.)

Now substituting the values for the individual voltages gives

Equation:

$$V = IR_1 + IR_2 + IR_3 = I(R_1 + R_2 + R_3).$$

Note that for the equivalent single series resistance R_s, we have

Equation:

$$V = IR_\mathrm{s}.$$

This implies that the total or equivalent series resistance R_s of three resistors is $R_\mathrm{s} = R_1 + R_2 + R_3$.

This logic is valid in general for any number of resistors in series; thus, the total resistance R_s of a series connection is

Equation:

$$R_\mathrm{s} = R_1 + R_2 + R_3 + \ldots,$$

as proposed. Since all of the current must pass through each resistor, it experiences the resistance of each, and resistances in series simply add up.

Example:

Calculating Resistance, Current, Voltage Drop, and Power Dissipation: Analysis of a Series Circuit

Suppose the voltage output of the battery in [link] is $12.0\ \mathrm{V}$, and the resistances are $R_1 = 1.00\ \Omega$, $R_2 = 6.00\ \Omega$, and $R_3 = 13.0\ \Omega$. (a) What is the total resistance? (b) Find the current. (c) Calculate the voltage drop in each resistor, and show these add to equal the voltage output of the source. (d) Calculate the power dissipated by each resistor. (e) Find the power output of the source, and show that it equals the total power dissipated by the resistors.

Strategy and Solution for (a)

The total resistance is simply the sum of the individual resistances, as given by this equation:

Equation:

$$\begin{aligned} R_{\text{s}} &= R_1 + R_2 + R_3 \\ &= 1.00\ \Omega + 6.00\ \Omega + 13.0\ \Omega \\ &= 20.0\ \Omega. \end{aligned}$$

Strategy and Solution for (b)

The current is found using Ohm's law, $V = \text{IR}$. Entering the value of the applied voltage and the total resistance yields the current for the circuit:

Equation:

$$I = \frac{V}{R_{\text{s}}} = \frac{12.0\ \text{V}}{20.0\ \Omega} = 0.600\ \text{A}.$$

Strategy and Solution for (c)

The voltage—or IR drop—in a resistor is given by Ohm's law. Entering the current and the value of the first resistance yields

Equation:

$$V_1 = IR_1 = (0.600\ \text{A})(1.0\ \Omega) = 0.600\ \text{V}.$$

Similarly,

Equation:

$$V_2 = IR_2 = (0.600\ \text{A})(6.0\ \Omega) = 3.60\ \text{V}$$

and

Equation:

$$V_3 = IR_3 = (0.600\ \text{A})(13.0\ \Omega) = 7.80\ \text{V}.$$

Discussion for (c)

The three IR drops add to $12.0\ \text{V}$, as predicted:

Equation:

$$V_1 + V_2 + V_3 = (0.600 + 3.60 + 7.80)\ \text{V} = 12.0\ \text{V}.$$

Strategy and Solution for (d)

The easiest way to calculate power in watts (W) dissipated by a resistor in a DC circuit is to use **Joule's law**, $P = \mathrm{IV}$, where P is electric power. In this case, each resistor has the same full current flowing through it. By substituting Ohm's law $V = \mathrm{IR}$ into Joule's law, we get the power dissipated by the first resistor as

Equation:

$$P_1 = I^2 R_1 = (0.600\ \mathrm{A})^2 (1.00\ \Omega) = 0.360\ \mathrm{W}.$$

Similarly,

Equation:

$$P_2 = I^2 R_2 = (0.600\ \mathrm{A})^2 (6.00\ \Omega) = 2.16\ \mathrm{W}$$

and

Equation:

$$P_3 = I^2 R_3 = (0.600\ \mathrm{A})^2 (13.0\ \Omega) = 4.68\ \mathrm{W}.$$

Discussion for (d)

Power can also be calculated using either $P = \mathrm{IV}$ or $P = \frac{V^2}{R}$, where V is the voltage drop across the resistor (not the full voltage of the source). The same values will be obtained.

Strategy and Solution for (e)

The easiest way to calculate power output of the source is to use $P = \mathrm{IV}$, where V is the source voltage. This gives

Equation:

$$P = (0.600\ \mathrm{A})(12.0\ \mathrm{V}) = 7.20\ \mathrm{W}.$$

Discussion for (e)

Note, coincidentally, that the total power dissipated by the resistors is also 7.20 W, the same as the power put out by the source. That is,

Equation:

$$P_1 + P_2 + P_3 = (0.360 + 2.16 + 4.68)\ \mathrm{W} = 7.20\ \mathrm{W}.$$

Power is energy per unit time (watts), and so conservation of energy requires the power output of the source to be equal to the total power dissipated by the resistors.

Note:

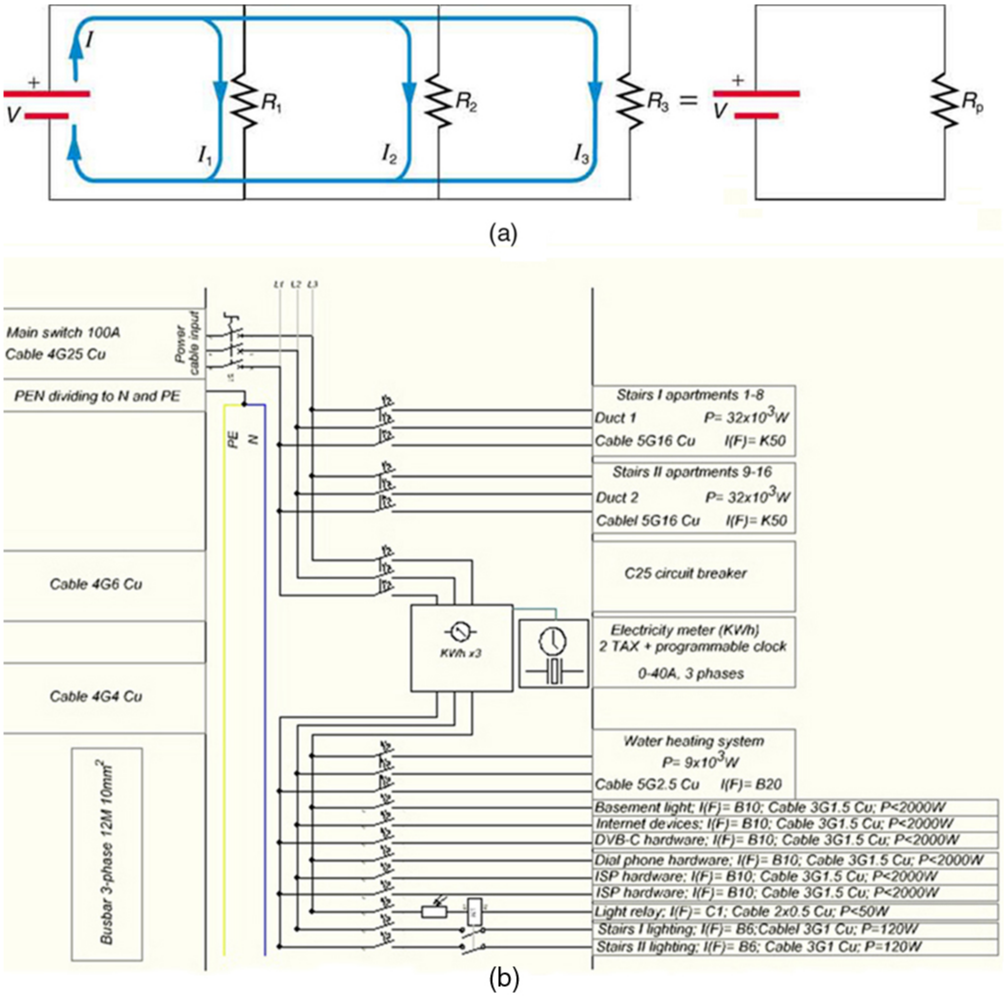
Major Features of Resistors in Series

1. Series resistances add: $R_s = R_1 + R_2 + R_3 + \ldots$.
2. The same current flows through each resistor in series.
3. Individual resistors in series do not get the total source voltage, but divide it.

Resistors in Parallel

[link] shows resistors in **parallel**, wired to a voltage source. Resistors are in parallel when each resistor is connected directly to the voltage source by connecting wires having negligible resistance. Each resistor thus has the full voltage of the source applied to it.

Each resistor draws the same current it would if it alone were connected to the voltage source (provided the voltage source is not overloaded). For example, an automobile’s headlights, radio, and so on, are wired in parallel, so that they utilize the full voltage of the source and can operate completely independently. The same is true in your house, or any building. (See [link] (b).)


I
+
V
R1
R2
R3
=
+
V
Rp
I1
I2
I3
(a)
L1 L2 L3
Main switch 100A
Cable 4G25 Cu
Power cable input
PEN dividing to N and PE
PE
N
Stairs I apartments 1-8
Duct 1 P= 32x10³W
Cable 5G16 Cu I(F)= K50
Stairs II apartments 9-16
Duct 2 P= 32x10³W
Cablel 5G16 Cu I(F)= K50
C25 circuit breaker
Cable 4G6 Cu
KVh x3
Electricity meter (KWh)
2 TAX + programmable clock
0-40A, 3 phases
Cable 4G4 Cu
Busbar 3-phase 12M 10mm²
Water heating system
P= 9x10³W
Cable 5G2.5 Cu I(F)= B20
Basement light; I(F)= B10; Cable 3G1.5 Cu; P<2000W
Internet devices; I(F)= B10; Cable 3G1.5 Cu; P<2000W
DVB-C hardware; I(F)= B10; Cable 3G1.5 Cu; P<2000W
Dial phone hardware; I(F)= B10; Cable 3G1.5 Cu; P<2000W
ISP hardware; I(F)= B10; Cable 3G1.5 Cu; P<2000W
ISP hardware; I(F)= B10; Cable 3G1.5 Cu; P<2000W
Light relay; I(F)= C1; Cable 2x0.5 Cu; P<50W
Stairs I lighting; I(F)= B6;Cablel 3G1 Cu; P=120W
Stairs II lighting; I(F)= B6; Cable 3G1 Cu; P=120W
(b)

(a) Three resistors connected in parallel to a battery and the equivalent single or parallel resistance. (b) Electrical power setup in a house. (credit: Dmitry G, Wikimedia Commons)

To find an expression for the equivalent parallel resistance R_{p}, let us consider the currents that flow and how they are related to resistance. Since each resistor in the circuit has the full voltage, the currents flowing through the individual resistors are $I_1 = \frac{V}{R_1}$, $I_2 = \frac{V}{R_2}$, and $I_3 = \frac{V}{R_3}$. Conservation of charge implies that the total current I produced by the source is the sum of these currents:

Equation:

$$I = I_1 + I_2 + I_3.$$

Substituting the expressions for the individual currents gives
Equation:

$$I = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3} = V\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right).$$

Note that Ohm’s law for the equivalent single resistance gives
Equation:

$$I = \frac{V}{R_{\text{p}}} = V\left(\frac{1}{R_{\text{p}}}\right).$$

The terms inside the parentheses in the last two equations must be equal. Generalizing to any number of resistors, the total resistance R_{p} of a parallel connection is related to the individual resistances by
Equation:

$$\frac{1}{R_{\text{p}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_{.3}} + \text{....}$$

This relationship results in a total resistance R_{p} that is less than the smallest of the individual resistances. (This is seen in the next example.) When resistors are connected in parallel, more current flows from the source than would flow for any of them individually, and so the total resistance is lower.

Example:
Calculating Resistance, Current, Power Dissipation, and Power Output: Analysis of a Parallel Circuit
Let the voltage output of the battery and resistances in the parallel connection in [link] be the same as the previously considered series

connection: $V = 12.0\text{ V}$, $R_1 = 1.00\ \Omega$, $R_2 = 6.00\ \Omega$, and $R_3 = 13.0\ \Omega$. (a) What is the total resistance? (b) Find the total current. (c) Calculate the currents in each resistor, and show these add to equal the total current output of the source. (d) Calculate the power dissipated by each resistor. (e) Find the power output of the source, and show that it equals the total power dissipated by the resistors.

Strategy and Solution for (a)

The total resistance for a parallel combination of resistors is found using the equation below. Entering known values gives

Equation:

$$\frac{1}{R_\text{p}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} = \frac{1}{1.00\ \Omega} + \frac{1}{6.00\ \Omega} + \frac{1}{13.0\ \Omega}.$$

Thus,

Equation:

$$\frac{1}{R_\text{p}} = \frac{1.00}{\Omega} + \frac{0.1667}{\Omega} + \frac{0.07692}{\Omega} = \frac{1.2436}{\Omega}.$$

(Note that in these calculations, each intermediate answer is shown with an extra digit.)

We must invert this to find the total resistance R_p. This yields

Equation:

$$R_\text{p} = \frac{1}{1.2436}\Omega = 0.8041\ \Omega.$$

The total resistance with the correct number of significant digits is $R_\text{p} = 0.804\ \Omega$.

Discussion for (a)

R_p is, as predicted, less than the smallest individual resistance.

Strategy and Solution for (b)

The total current can be found from Ohm’s law, substituting R_p for the total resistance. This gives

Equation:

$$I = \frac{V}{R_\mathrm{p}} = \frac{12.0\ \mathrm{V}}{0.8041\ \Omega} = 14.92\ \mathrm{A}.$$

Discussion for (b)
Current I for each device is much larger than for the same devices connected in series (see the previous example). A circuit with parallel connections has a smaller total resistance than the resistors connected in series.

Strategy and Solution for (c)
The individual currents are easily calculated from Ohm’s law, since each resistor gets the full voltage. Thus,

Equation:

$$I_1 = \frac{V}{R_1} = \frac{12.0\ \mathrm{V}}{1.00\,\Omega} = 12.0\ \mathrm{A}.$$

Similarly,

Equation:

$$I_2 = \frac{V}{R_2} = \frac{12.0\ \mathrm{V}}{6.00\,\Omega} = 2.00\ \mathrm{A}$$

and

Equation:

$$I_3 = \frac{V}{R_3} = \frac{12.0\ \mathrm{V}}{13.0\,\Omega} = 0.92\ \mathrm{A}.$$

Discussion for (c)
The total current is the sum of the individual currents:

Equation:

$$I_1 + I_2 + I_3 = 14.92\ \mathrm{A}.$$

This is consistent with conservation of charge.

Strategy and Solution for (d)
The power dissipated by each resistor can be found using any of the equations relating power to current, voltage, and resistance, since all three

are known. Let us use $P=\frac{V^2}{R}$, since each resistor gets full voltage. Thus,

Equation:

$$P_1=\frac{V^2}{R_1}=\frac{(12.0\text{ V})^2}{1.00\ \Omega}=144\text{ W}.$$

Similarly,

Equation:

$$P_2=\frac{V^2}{R_2}=\frac{(12.0\text{ V})^2}{6.00\ \Omega}=24.0\text{ W}$$

and

Equation:

$$P_3=\frac{V^2}{R_3}=\frac{(12.0\text{ V})^2}{13.0\ \Omega}=11.1\text{ W}.$$

Discussion for (d)

The power dissipated by each resistor is considerably higher in parallel than when connected in series to the same voltage source.

Strategy and Solution for (e)

The total power can also be calculated in several ways. Choosing $P=\text{IV}$, and entering the total current, yields

Equation:

$$P=\text{IV}=(14.92\text{ A})(12.0\text{ V})=179\text{ W}.$$

Discussion for (e)

Total power dissipated by the resistors is also 179 W:

Equation:

$$P_1+P_2+P_3=144\text{ W}+24.0\text{ W}+11.1\text{ W}=179\text{ W}.$$

This is consistent with the law of conservation of energy.

Overall Discussion

Note that both the currents and powers in parallel connections are greater than for the same devices in series.

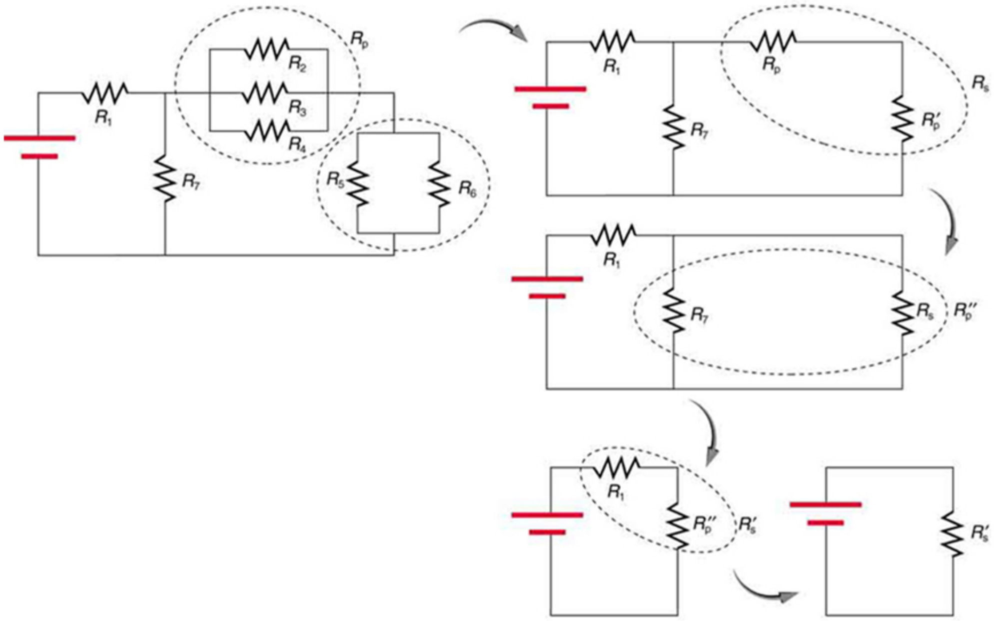
Note:
Major Features of Resistors in Parallel

1. Parallel resistance is found from $\frac{1}{R_\text{p}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + ...$, and it is smaller than any individual resistance in the combination.
2. Each resistor in parallel has the same full voltage of the source applied to it. (Power distribution systems most often use parallel connections to supply the myriad devices served with the same voltage and to allow them to operate independently.)
3. Parallel resistors do not each get the total current; they divide it.

Combinations of Series and Parallel

More complex connections of resistors are sometimes just combinations of series and parallel. These are commonly encountered, especially when wire resistance is considered. In that case, wire resistance is in series with other resistances that are in parallel.

Combinations of series and parallel can be reduced to a single equivalent resistance using the technique illustrated in [link]. Various parts are identified as either series or parallel, reduced to their equivalents, and further reduced until a single resistance is left. The process is more time consuming than difficult.



This combination of seven resistors has both series and parallel parts. Each is identified and reduced to an equivalent resistance, and these are further reduced until a single equivalent resistance is reached.

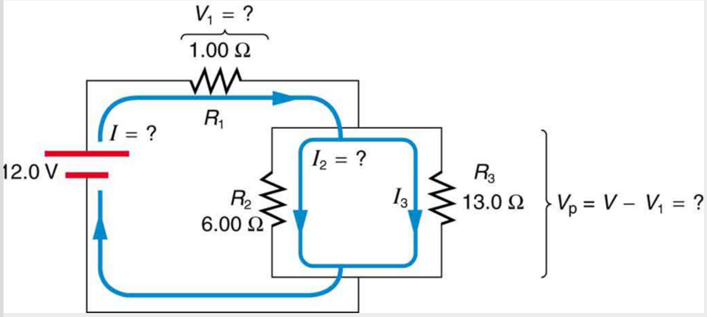
The simplest combination of series and parallel resistance, shown in [link], is also the most instructive, since it is found in many applications. For example, R_1 could be the resistance of wires from a car battery to its electrical devices, which are in parallel. R_2 and R_3 could be the starter motor and a passenger compartment light. We have previously assumed that wire resistance is negligible, but, when it is not, it has important effects, as the next example indicates.

Example:

Calculating Resistance, IR Drop, Current, and Power Dissipation: Combining Series and Parallel Circuits

[link] shows the resistors from the previous two examples wired in a different way—a combination of series and parallel. We can consider R_1 to be the resistance of wires leading to R_2 and R_3. (a) Find the total

resistance. (b) What is the IR drop in R_1? (c) Find the current I_2 through R_2. (d) What power is dissipated by R_2?



These three resistors are connected to a voltage source so that R_2 and R_3 are in parallel with one another and that combination is in series with R_1.

Strategy and Solution for (a)

To find the total resistance, we note that R_2 and R_3 are in parallel and their combination R_p is in series with R_1. Thus the total (equivalent) resistance of this combination is

Equation:

$$R_\mathrm{tot} = R_1 + R_\mathrm{p}.$$

First, we find R_p using the equation for resistors in parallel and entering known values:

Equation:

$$\frac{1}{R_\mathrm{p}} = \frac{1}{R_2} + \frac{1}{R_3} = \frac{1}{6.00\ \Omega} + \frac{1}{13.0\ \Omega} = \frac{0.2436}{\Omega}.$$

Inverting gives

Equation:

$$R_\mathrm{p} = \frac{1}{0.2436}\Omega = 4.11\ \Omega.$$

So the total resistance is

Equation:

$$R_{\text{tot}} = R_1 + R_{\text{p}} = 1.00\ \Omega + 4.11\ \Omega = 5.11\ \Omega.$$

Discussion for (a)

The total resistance of this combination is intermediate between the pure series and pure parallel values ($20.0\ \Omega$ and $0.804\ \Omega$, respectively) found for the same resistors in the two previous examples.

Strategy and Solution for (b)

To find the IR drop in R_1, we note that the full current I flows through R_1. Thus its IR drop is

Equation:

$$V_1 = \text{IR}_1.$$

We must find I before we can calculate V_1. The total current I is found using Ohm’s law for the circuit. That is,

Equation:

$$I = \frac{V}{R_{\text{tot}}} = \frac{12.0\ \text{V}}{5.11\ \Omega} = 2.35\ \text{A}.$$

Entering this into the expression above, we get

Equation:

$$V_1 = \text{IR}_1 = (2.35\ \text{A})(1.00\ \Omega) = 2.35\ \text{V}.$$

Discussion for (b)

The voltage applied to R_2 and R_3 is less than the total voltage by an amount V_1. When wire resistance is large, it can significantly affect the operation of the devices represented by R_2 and R_3.

Strategy and Solution for (c)

To find the current through R_2, we must first find the voltage applied to it. We call this voltage V_{p}, because it is applied to a parallel combination of resistors. The voltage applied to both R_2 and R_3 is reduced by the amount V_1, and so it is

Equation:

$$V_p = V - V_1 = 12.0\ \text{V} - 2.35\ \text{V} = 9.65\ \text{V}.$$

Now the current I_2 through resistance R_2 is found using Ohm’s law:
Equation:

$$I_2 = \frac{V_p}{R_2} = \frac{9.65\ \text{V}}{6.00\ \Omega} = 1.61\ \text{A}.$$

Discussion for (c)
The current is less than the 2.00 A that flowed through R_2 when it was connected in parallel to the battery in the previous parallel circuit example.
Strategy and Solution for (d)
The power dissipated by R_2 is given by
Equation:

$$P_2 = (I_2)^2 R_2 = (1.61\ \text{A})^2 (6.00\ \Omega) = 15.5\ \text{W}.$$

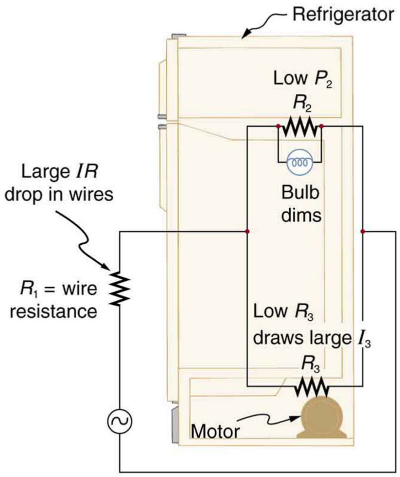
Discussion for (d)
The power is less than the 24.0 W this resistor dissipated when connected in parallel to the 12.0-V source.

Practical Implications

One implication of this last example is that resistance in wires reduces the current and power delivered to a resistor. If wire resistance is relatively large, as in a worn (or a very long) extension cord, then this loss can be significant. If a large current is drawn, the IR drop in the wires can also be significant.

For example, when you are rummaging in the refrigerator and the motor comes on, the refrigerator light dims momentarily. Similarly, you can see the passenger compartment light dim when you start the engine of your car (although this may be due to resistance inside the battery itself).

What is happening in these high-current situations is illustrated in [link]. The device represented by R_3 has a very low resistance, and so when it is switched on, a large current flows. This increased current causes a larger IR drop in the wires represented by R_1, reducing the voltage across the light bulb (which is R_2), which then dims noticeably.


Refrigerator
Low P_2
R_2
Bulb dims
Large IR drop in wires
R_1 = wire resistance
Low R_3 draws large I_3
R_3
Motor

Why do lights dim when a large appliance is switched on? The answer is that the large current the appliance motor draws causes a significant IR drop in the wires and reduces the voltage across the light.

Exercise:
Check Your Understanding

Problem:

Can any arbitrary combination of resistors be broken down into series and parallel combinations? See if you can draw a circuit diagram of resistors that cannot be broken down into combinations of series and parallel.

Solution:

No, there are many ways to connect resistors that are not combinations of series and parallel, including loops and junctions. In such cases Kirchhoff’s rules, to be introduced in Kirchhoff’s Rules, will allow you to analyze the circuit.

Note:

Problem-Solving Strategies for Series and Parallel Resistors

1. Draw a clear circuit diagram, labeling all resistors and voltage sources. This step includes a list of the knowns for the problem, since they are labeled in your circuit diagram.
2. Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful.
3. Determine whether resistors are in series, parallel, or a combination of both series and parallel. Examine the circuit diagram to make this assessment. Resistors are in series if the same current must pass sequentially through them.
4. Use the appropriate list of major features for series or parallel connections to solve for the unknowns. There is one list for series and another for parallel. If your problem has a combination of series and parallel, reduce it in steps by considering individual groups of series or parallel connections, as done in this module and the examples. Special note: When finding R_{p}, the reciprocal must be taken with care.
5. Check to see whether the answers are reasonable and consistent. Units and numerical results must be reasonable. Total series resistance

should be greater, whereas total parallel resistance should be smaller, for example. Power should be greater for the same devices in parallel compared with series, and so on.

Section Summary

- The total resistance of an electrical circuit with resistors wired in a series is the sum of the individual resistances: $R_\text{s} = R_1 + R_2 + R_3 + ...$
- Each resistor in a series circuit has the same amount of current flowing through it.
- The voltage drop, or power dissipation, across each individual resistor in a series is different, and their combined total adds up to the power source input.
- The total resistance of an electrical circuit with resistors wired in parallel is less than the lowest resistance of any of the components and can be determined using the formula:
  Equation:

  $$\frac{1}{R_\text{p}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + ...$$

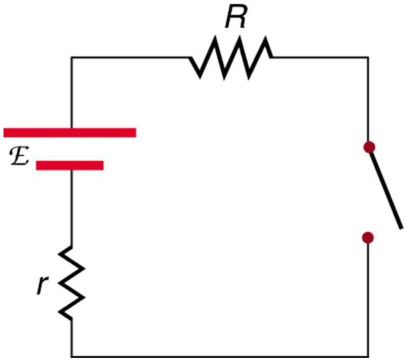
- Each resistor in a parallel circuit has the same full voltage of the source applied to it.
- The current flowing through each resistor in a parallel circuit is different, depending on the resistance.
- If a more complex connection of resistors is a combination of series and parallel, it can be reduced to a single equivalent resistance by identifying its various parts as series or parallel, reducing each to its equivalent, and continuing until a single resistance is eventually reached.

Conceptual Questions

Exercise:

Problem:

A switch has a variable resistance that is nearly zero when closed and extremely large when open, and it is placed in series with the device it controls. Explain the effect the switch in [link] has on current when open and when closed.



A switch is ordinarily in series with a resistance and voltage source. Ideally, the switch has nearly zero resistance when closed but has an extremely large resistance when open. (Note that in this diagram, the script E represents the voltage (or electromotive force) of the battery.)

Exercise:

Problem: What is the voltage across the open switch in [link]?

Exercise:

Problem:

There is a voltage across an open switch, such as in [link]. Why, then, is the power dissipated by the open switch small?

Exercise:

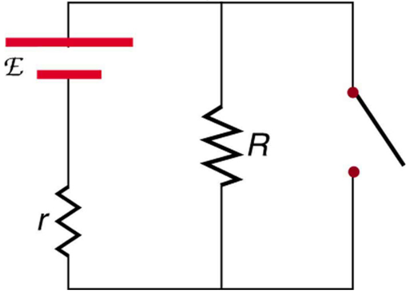
Problem:

Why is the power dissipated by a closed switch, such as in [link], small?

Exercise:

Problem:

A student in a physics lab mistakenly wired a light bulb, battery, and switch as shown in [link]. Explain why the bulb is on when the switch is open, and off when the switch is closed. (Do not try this—it is hard on the battery!)



A wiring mistake put this switch in parallel with the device represented by R. (Note that in this diagram, the script E represents the voltage (or

electromotive
force) of the
battery.)

Exercise:

Problem:

Knowing that the severity of a shock depends on the magnitude of the current through your body, would you prefer to be in series or parallel with a resistance, such as the heating element of a toaster, if shocked by it? Explain.

Exercise:

Problem:

Would your headlights dim when you start your car’s engine if the wires in your automobile were superconductors? (Do not neglect the battery’s internal resistance.) Explain.

Exercise:

Problem:

Some strings of holiday lights are wired in series to save wiring costs. An old version utilized bulbs that break the electrical connection, like an open switch, when they burn out. If one such bulb burns out, what happens to the others? If such a string operates on 120 V and has 40 identical bulbs, what is the normal operating voltage of each? Newer versions use bulbs that short circuit, like a closed switch, when they burn out. If one such bulb burns out, what happens to the others? If such a string operates on 120 V and has 39 remaining identical bulbs, what is then the operating voltage of each?

Exercise:

Problem:

If two household lightbulbs rated 60 W and 100 W are connected in series to household power, which will be brighter? Explain.

Exercise:

Problem:

Suppose you are doing a physics lab that asks you to put a resistor into a circuit, but all the resistors supplied have a larger resistance than the requested value. How would you connect the available resistances to attempt to get the smaller value asked for?

Exercise:

Problem:

Before World War II, some radios got power through a “resistance cord” that had a significant resistance. Such a resistance cord reduces the voltage to a desired level for the radio’s tubes and the like, and it saves the expense of a transformer. Explain why resistance cords become warm and waste energy when the radio is on.

Exercise:

Problem:

Some light bulbs have three power settings (not including zero), obtained from multiple filaments that are individually switched and wired in parallel. What is the minimum number of filaments needed for three power settings?

Problem Exercises

Note: Data taken from figures can be assumed to be accurate to three significant digits.

Exercise:

Problem:

(a) What is the resistance of ten $275\text{-}\Omega$ resistors connected in series? (b) In parallel?

Solution:

(a) $2.75\text{ k}\Omega$

(b) $27.5\ \Omega$

Exercise:

Problem:

(a) What is the resistance of a 1.00×10^2 $\text{-}\Omega$, a $2.50\text{-k}\Omega$, and a $4.00\text{-k}\Omega$ resistor connected in series? (b) In parallel?

Exercise:

Problem:

What are the largest and smallest resistances you can obtain by connecting a $36.0\text{-}\Omega$, a $50.0\text{-}\Omega$, and a $700\text{-}\Omega$ resistor together?

Solution:

(a) $786\ \Omega$

(b) $20.3\ \Omega$

Exercise:

Problem:

An 1800-W toaster, a 1400-W electric frying pan, and a 75-W lamp are plugged into the same outlet in a 15-A, 120-V circuit. (The three devices are in parallel when plugged into the same socket.). (a) What current is drawn by each device? (b) Will this combination blow the 15-A fuse?

Exercise:

Problem:

Your car’s 30.0-W headlight and 2.40-kW starter are ordinarily connected in parallel in a 12.0-V system. What power would one headlight and the starter consume if connected in series to a 12.0-V battery? (Neglect any other resistance in the circuit and any change in resistance in the two devices.)

Solution:

29.6 W

Exercise:

Problem:

(a) Given a 48.0-V battery and $24.0\text{-}\Omega$ and $96.0\text{-}\Omega$ resistors, find the current and power for each when connected in series. (b) Repeat when the resistances are in parallel.

Exercise:

Problem:

Referring to the example combining series and parallel circuits and [link], calculate I_3 in the following two different ways: (a) from the known values of I and I_2; (b) using Ohm’s law for R_3. In both parts explicitly show how you follow the steps in the Problem-Solving Strategies for Series and Parallel Resistors.

Solution:

(a) 0.74 A

(b) 0.742 A

Exercise:

Problem:

Referring to [link]: (a) Calculate P_3 and note how it compares with P_3 found in the first two example problems in this module. (b) Find the total power supplied by the source and compare it with the sum of the powers dissipated by the resistors.

Exercise:

Problem:

Refer to [link] and the discussion of lights dimming when a heavy appliance comes on. (a) Given the voltage source is 120 V, the wire resistance is $0.400\ \Omega$, and the bulb is nominally 75.0 W, what power will the bulb dissipate if a total of 15.0 A passes through the wires when the motor comes on? Assume negligible change in bulb resistance. (b) What power is consumed by the motor?

Solution:

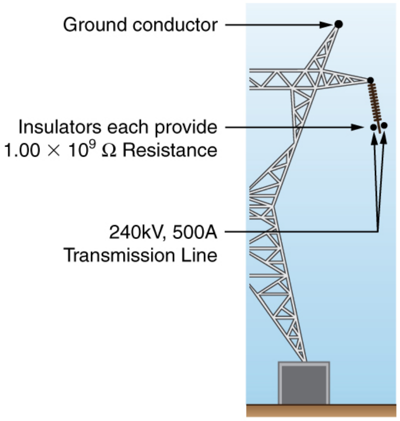
(a) 60.8 W

(b) 3.18 kW

Exercise:

Problem:

A 240-kV power transmission line carrying 5.00×10^2 A is hung from grounded metal towers by ceramic insulators, each having a 1.00×10^9 -Ω resistance. [link]. (a) What is the resistance to ground of 100 of these insulators? (b) Calculate the power dissipated by 100 of them. (c) What fraction of the power carried by the line is this? Explicitly show how you follow the steps in the Problem-Solving Strategies for Series and Parallel Resistors.


Ground conductor
Insulators each provide
1.00 × 10^9 Ω Resistance
240kV, 500A
Transmission Line

High-voltage (240-kV) transmission line carrying 5.00×10^2 A is hung from a grounded metal transmission tower. The row of ceramic insulators provide $1.00 \times 10^9 \Omega$ of resistance each.

Exercise:

Problem:

Show that if two resistors R_1 and R_2 are combined and one is much greater than the other ($R_1 >> R_2$): (a) Their series resistance is very nearly equal to the greater resistance R_1. (b) Their parallel resistance is very nearly equal to smaller resistance R_2.

Solution:

(a) $$\begin{array}{c} R_\mathrm{s} = R_1 + R_2 \\ \Rightarrow R_\mathrm{s} \approx R_1 (R_1 >> R_2) \end{array}$$

(b) $\frac{1}{R_{\text{p}}} = \frac{1}{R_1} + \frac{1}{R_2} = \frac{R_1+R_2}{R_1R_2}$,

so that

$$R_{\text{p}} = \frac{R_1R_2}{R_1+R_2} \approx \frac{R_1R_2}{R_1} = R_2(R_1 >> R_2).$$

Exercise:

Problem: **Unreasonable Results**

Two resistors, one having a resistance of $145\ \Omega$, are connected in parallel to produce a total resistance of $150\ \Omega$. (a) What is the value of the second resistance? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem: **Unreasonable Results**

Two resistors, one having a resistance of $900\ \text{k}\Omega$, are connected in series to produce a total resistance of $0.500\ \text{M}\Omega$. (a) What is the value of the second resistance? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $-400\ \text{k}\Omega$

(b) Resistance cannot be negative.

(c) Series resistance is said to be less than one of the resistors, but it must be greater than any of the resistors.

Glossary

series

a sequence of resistors or other components wired into a circuit one after the other

resistor
a component that provides resistance to the current flowing through an electrical circuit

resistance
causing a loss of electrical power in a circuit

Ohm’s law
the relationship between current, voltage, and resistance within an electrical circuit: $V = \mathrm{IR}$

voltage
the electrical potential energy per unit charge; electric pressure created by a power source, such as a battery

voltage drop
the loss of electrical power as a current travels through a resistor, wire or other component

current
the flow of charge through an electric circuit past a given point of measurement

Joule’s law
the relationship between potential electrical power, voltage, and resistance in an electrical circuit, given by: $P_e = \mathrm{IV}$

parallel
the wiring of resistors or other components in an electrical circuit such that each component receives an equal voltage from the power source; often pictured in a ladder-shaped diagram, with each component on a rung of the ladder

Electromotive Force: Terminal Voltage

- Compare and contrast the voltage and the electromagnetic force of an electric power source.
- Describe what happens to the terminal voltage, current, and power delivered to a load as internal resistance of the voltage source increases (due to aging of batteries, for example).
- Explain why it is beneficial to use more than one voltage source connected in parallel.

When you forget to turn off your car lights, they slowly dim as the battery runs down. Why don't they simply blink off when the battery's energy is gone? Their gradual dimming implies that battery output voltage decreases as the battery is depleted.

Furthermore, if you connect an excessive number of 12-V lights in parallel to a car battery, they will be dim even when the battery is fresh and even if the wires to the lights have very low resistance. This implies that the battery's output voltage is reduced by the overload.

The reason for the decrease in output voltage for depleted or overloaded batteries is that all voltage sources have two fundamental parts—a source of electrical energy and an **internal resistance**. Let us examine both.

Electromotive Force

You can think of many different types of voltage sources. Batteries themselves come in many varieties. There are many types of mechanical/electrical generators, driven by many different energy sources, ranging from nuclear to wind. Solar cells create voltages directly from light, while thermoelectric devices create voltage from temperature differences.

A few voltage sources are shown in [link]. All such devices create a **potential difference** and can supply current if connected to a resistance. On the small scale, the potential difference creates an electric field that exerts force on charges, causing current. We thus use the name **electromotive force**, abbreviated emf.

Emf is not a force at all; it is a special type of potential difference. To be precise, the electromotive force (emf) is the potential difference of a source when no current is flowing. Units of emf are volts.

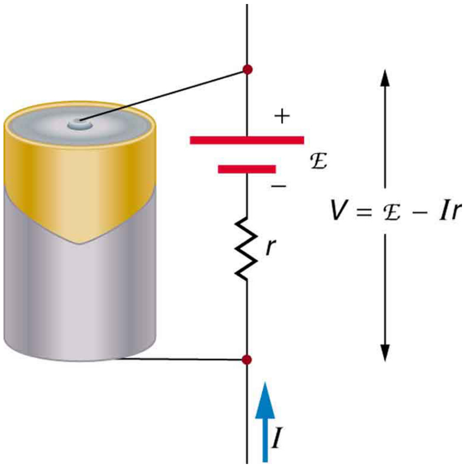
A variety of voltage sources (clockwise from top left): the Brazos Wind Farm in Fluvanna, Texas (credit: Leaflet, Wikimedia Commons); the Krasnoyarsk Dam in Russia (credit: Alex Polezhaev); a solar farm (credit: U.S. Department of Energy); and a group of nickel metal hydride batteries (credit: Tiaa Monto). The voltage output of each depends on its construction and load, and equals emf only if there is no load.

Electromotive force is directly related to the source of potential difference, such as the particular combination of chemicals in a battery. However, emf differs from the voltage output of the device when current flows. The voltage across the terminals of a battery, for example, is less than the emf when the battery supplies current, and it declines further as the battery is depleted or loaded down. However, if the device's output voltage can be measured without drawing current, then output voltage will equal emf (even for a very depleted battery).

Internal Resistance

As noted before, a 12-V truck battery is physically larger, contains more charge and energy, and can deliver a larger current than a 12-V motorcycle battery. Both are lead-acid batteries with identical emf, but, because of its size, the truck battery has a smaller internal resistance r. Internal resistance is the inherent resistance to the flow of current within the source itself.

[link] is a schematic representation of the two fundamental parts of any voltage source. The emf (represented by a script E in the figure) and internal resistance r are in series. The smaller the internal resistance for a given emf, the more current and the more power the source can supply.



Any voltage source (in this case, a carbon-zinc dry cell) has an emf related to its source of potential difference, and an internal resistance r related to its construction. (Note that the script E stands for emf.). Also shown are the output terminals across which the terminal voltage V is measured. Since $V = \text{emf} - \text{Ir}$, terminal voltage equals emf only if there is no current flowing.

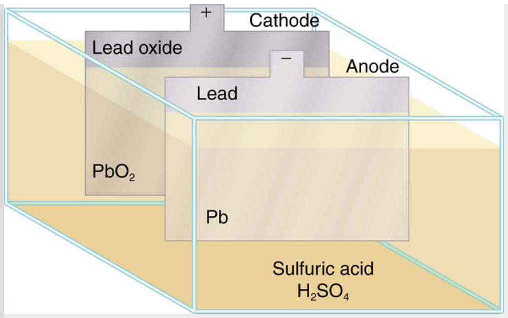
The internal resistance r can behave in complex ways. As noted, r increases as a battery is depleted. But internal resistance may also depend on the magnitude and direction of the current through a voltage source, its temperature, and even its history. The internal resistance of rechargeable nickel-cadmium cells, for example, depends on how many times and how deeply they have been depleted.

Note:

Things Great and Small: The Submicroscopic Origin of Battery Potential

Various types of batteries are available, with emfs determined by the combination of chemicals involved. We can view this as a molecular reaction (what much of chemistry is about) that separates charge.

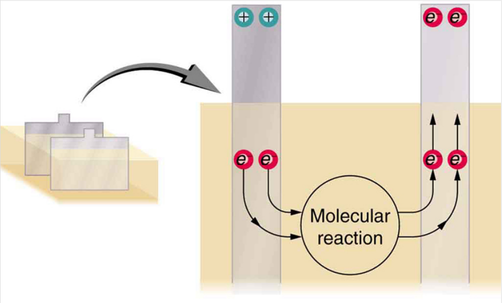
The lead-acid battery used in cars and other vehicles is one of the most common types. A single cell (one of six) of this battery is seen in [link]. The cathode (positive) terminal of the cell is connected to a lead oxide plate, while the anode (negative) terminal is connected to a lead plate. Both plates are immersed in sulfuric acid, the electrolyte for the system.



Artist’s conception of a lead-acid cell. Chemical reactions in a lead-acid cell separate charge, sending negative charge to the anode, which is connected to the lead plates. The lead oxide plates are connected to the positive or cathode terminal of the cell. Sulfuric acid conducts the charge as well as participating in the chemical reaction.

The details of the chemical reaction are left to the reader to pursue in a chemistry text, but their results at the molecular level help explain the potential created by the battery. [link] shows the result of a single chemical reaction. Two electrons are placed on the anode, making it negative, provided that the cathode supplied two electrons. This leaves the cathode positively charged, because it has lost two electrons. In short, a separation of charge has been driven by a chemical reaction.

Note that the reaction will not take place unless there is a complete circuit to allow two electrons to be supplied to the cathode. Under many circumstances, these electrons come from the anode, flow through a resistance, and return to the cathode. Note also that since the chemical reactions involve substances with resistance, it is not possible to create the emf without an internal resistance.



Artist’s conception of two electrons being forced onto the anode of a cell and two electrons being removed from the cathode of the cell. The chemical reaction in a lead-acid battery places two electrons on the anode and removes two from the cathode. It requires a closed circuit to proceed, since the two electrons must be supplied to the cathode.

Why are the chemicals able to produce a unique potential difference? Quantum mechanical descriptions of molecules, which take into account the types of atoms and numbers of electrons in them, are able to predict the energy states they can have and the energies of reactions between them.

In the case of a lead-acid battery, an energy of 2 eV is given to each electron sent to the anode. Voltage is defined as the electrical potential energy divided by charge: $V=\frac{P_{\text{E}}}{q}$. An electron volt is the energy given to a single electron by a voltage of 1 V. So the voltage here is 2 V, since 2 eV is given to each electron. It is the energy produced in each molecular reaction that produces the voltage. A different reaction produces a different energy and, hence, a different voltage.

Terminal Voltage

The voltage output of a device is measured across its terminals and, thus, is called its **terminal voltage** V. Terminal voltage is given by
Equation:

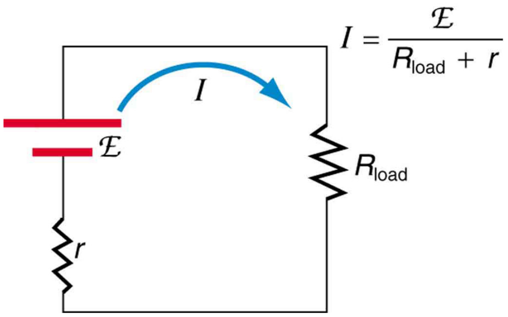
$$V=\text{emf}-\text{Ir},$$

where r is the internal resistance and I is the current flowing at the time of the measurement.

I is positive if current flows away from the positive terminal, as shown in [link]. You can see that the larger the current, the smaller the terminal voltage. And it is likewise true that the larger the internal resistance, the smaller the terminal voltage.

Suppose a load resistance R_{load} is connected to a voltage source, as in [link]. Since the resistances are in series, the total resistance in the circuit is $R_{\text{load}}+r$. Thus the current is given by Ohm’s law to be
Equation:

$$I=\frac{\text{emf}}{R_{\text{load}}+r}.$$



Schematic of a voltage source and its load R_{load}. Since the internal resistance r is in series with the load, it can significantly affect the terminal voltage and current delivered to the load. (Note that the script E stands for emf.)

We see from this expression that the smaller the internal resistance r, the greater the current the voltage source supplies to its load R_{load}. As batteries are depleted, r increases. If r becomes a significant fraction of the load resistance, then the current is significantly reduced, as the following example illustrates.

Example:

Calculating Terminal Voltage, Power Dissipation, Current, and Resistance: Terminal Voltage and Load

A certain battery has a 12.0-V emf and an internal resistance of 0.100 Ω. (a) Calculate its terminal voltage when connected to a 10.0-Ω load. (b) What is the terminal voltage when connected to a 0.500-Ω load? (c) What power does the 0.500-Ω load dissipate? (d) If the internal resistance grows to 0.500 Ω, find the current, terminal voltage, and power dissipated by a 0.500-Ω load.

Strategy

The analysis above gave an expression for current when internal resistance is taken into account. Once the current is found, the terminal voltage can be calculated using the equation $V = \text{emf} - \text{Ir}$. Once current is found, the power dissipated by a resistor can also be found.

Solution for (a)

Entering the given values for the emf, load resistance, and internal resistance into the expression above yields

Equation:

$$I = \frac{\text{emf}}{R_{\text{load}} + r} = \frac{12.0\text{ V}}{10.1\,\Omega} = 1.188\text{ A}.$$

Enter the known values into the equation $V = \text{emf} - \text{Ir}$ to get the terminal voltage:

Equation:

$$\begin{aligned} V &= \text{emf} - Ir = 12.0\text{ V} - (1.188\text{ A})(0.100\,\Omega) \\ &= 11.9\text{ V}. \end{aligned}$$

Discussion for (a)

The terminal voltage here is only slightly lower than the emf, implying that 10.0 Ω is a light load for this particular battery.

Solution for (b)

Similarly, with $R_{\text{load}} = 0.500\,\Omega$, the current is

Equation:

$$I = \frac{\text{emf}}{R_{\text{load}} + r} = \frac{12.0\text{ V}}{0.600\,\Omega} = 20.0\text{ A}.$$

The terminal voltage is now

Equation:

$$\begin{aligned} V &= \text{emf} - Ir = 12.0\text{ V} - (20.0\text{ A})(0.100\,\Omega) \\ &= 10.0\text{ V}. \end{aligned}$$

Discussion for (b)

This terminal voltage exhibits a more significant reduction compared with emf, implying $0.500\,\Omega$ is a heavy load for this battery.

Solution for (c)

The power dissipated by the 0.500 - Ω load can be found using the formula $P = I^2R$. Entering the known values gives

Equation:

$$P_{\text{load}} = I^2R_{\text{load}} = (20.0\text{ A})^2(0.500\ \Omega) = 2.00 \times 10^2\text{ W}.$$

Discussion for (c)

Note that this power can also be obtained using the expressions $\frac{V^2}{R}$ or IV, where V is the terminal voltage (10.0 V in this case).

Solution for (d)

Here the internal resistance has increased, perhaps due to the depletion of the battery, to the point where it is as great as the load resistance. As before, we first find the current by entering the known values into the expression, yielding

Equation:

$$I = \frac{\text{emf}}{R_{\text{load}} + r} = \frac{12.0\text{ V}}{1.00\ \Omega} = 12.0\text{ A}.$$

Now the terminal voltage is

Equation:

$$\begin{aligned} V &= \text{emf} - Ir = 12.0\text{ V} - (12.0\text{ A})(0.500\ \Omega) \\ &= 6.00\text{ V}, \end{aligned}$$

and the power dissipated by the load is

Equation:

$$P_{\text{load}} = I^2R_{\text{load}} = (12.0\text{ A})^2(0.500\ \Omega) = 72.0\text{ W}.$$

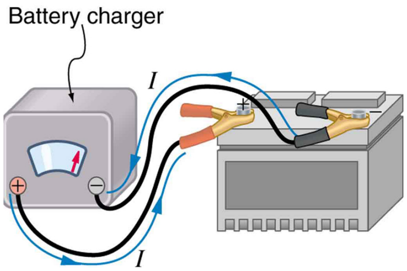
Discussion for (d)

We see that the increased internal resistance has significantly decreased terminal voltage, current, and power delivered to a load.

Battery testers, such as those in [link], use small load resistors to intentionally draw current to determine whether the terminal voltage drops below an acceptable level. They really test the internal resistance of the battery. If internal resistance is high, the battery is weak, as evidenced by its low terminal voltage.

These two battery testers measure terminal voltage under a load to determine the condition of a battery. The large device is being used by a U.S. Navy electronics technician to test large batteries aboard the aircraft carrier USS *Nimitz* and has a small resistance that can dissipate large amounts of power. (credit: U.S. Navy photo by Photographer’s Mate Airman Jason A. Johnston) The small device is used on small batteries and has a digital display to indicate the acceptability of their terminal voltage. (credit: Keith Williamson)

Some batteries can be recharged by passing a current through them in the direction opposite to the current they supply to a resistance. This is done routinely in cars and batteries for small electrical appliances and electronic devices, and is represented pictorially in [link]. The voltage output of the battery charger must be greater than the emf of the battery to reverse current through it. This will cause the terminal voltage of the battery to be greater than the emf, since $V = \text{emf} - \text{Ir}$, and I is now negative.



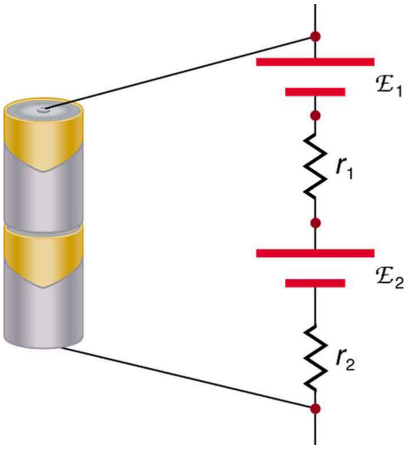
A car battery charger reverses the normal direction of current through a battery, reversing its chemical reaction and replenishing its chemical potential.

Multiple Voltage Sources

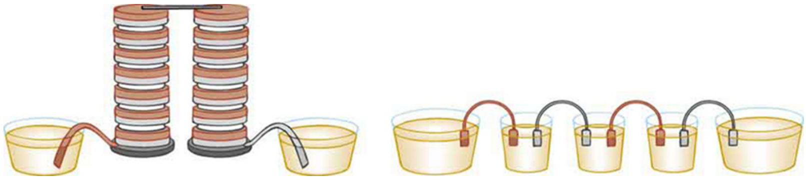
There are two voltage sources when a battery charger is used. Voltage sources connected in series are relatively simple. When voltage sources are in series, their internal resistances add and their emfs add algebraically. (See [link].) Series connections of voltage sources are common—for example, in flashlights, toys, and other appliances. Usually, the cells are in series in order to produce a larger total emf.

But if the cells oppose one another, such as when one is put into an appliance backward, the total emf is less, since it is the algebraic sum of the individual emfs.

A battery is a multiple connection of voltaic cells, as shown in [link]. The disadvantage of series connections of cells is that their internal resistances add. One of the authors once owned a 1957 MGA that had two 6-V batteries in series, rather than a single 12-V battery. This arrangement produced a large internal resistance that caused him many problems in starting the engine.

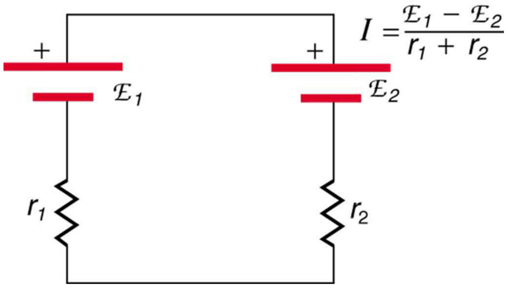


A series connection of two voltage sources. The emfs (each labeled with a script E) and internal resistances add, giving a total emf of $\text{emf}_1 + \text{emf}_2$ and a total internal resistance of $r_1 + r_2$.

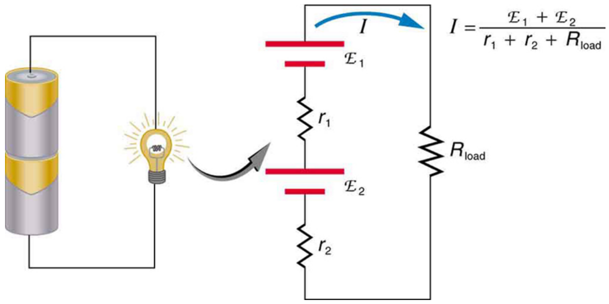


Batteries are multiple connections of individual cells, as shown in this modern rendition of an old print. Single cells, such as AA or C cells, are commonly called batteries, although this is technically incorrect.

If the *series* connection of two voltage sources is made into a complete circuit with the emfs in opposition, then a current of magnitude $I=\frac{(\text{emf}_1-\text{emf}_2)}{r_1+r_2}$ flows. See [link], for example, which shows a circuit exactly analogous to the battery charger discussed above. If two voltage sources in series with emfs in the same sense are connected to a load R_{load}, as in [link], then $I=\frac{(\text{emf}_1+\text{emf}_2)}{r_1+r_2+R_{\text{load}}}$ flows.



These two voltage sources are connected in series with their emfs in opposition. Current flows in the direction of the greater emf and is limited to $I = \frac{(\mathrm{emf}_1 - \mathrm{emf}_2)}{r_1 + r_2}$ by the sum of the internal resistances. (Note that each emf is represented by script E in the figure.) A battery charger connected to a battery is an example of such a connection. The charger must have a larger emf than the battery to reverse current through it.



I
$I = \frac{\mathcal{E}_1 + \mathcal{E}_2}{r_1 + r_2 + R_{\mathrm{load}}}$
$\mathcal{E}_1$
r_1
R_{load}
$\mathcal{E}_2$
r_2

This schematic represents a flashlight with two cells (voltage sources) and a single bulb (load resistance) in series. The current that flows is $I = \frac{(\mathrm{emf}_1 + \mathrm{emf}_2)}{r_1 + r_2 + R_{\mathrm{load}}}$. (Note that each emf is represented by script E in the figure.)

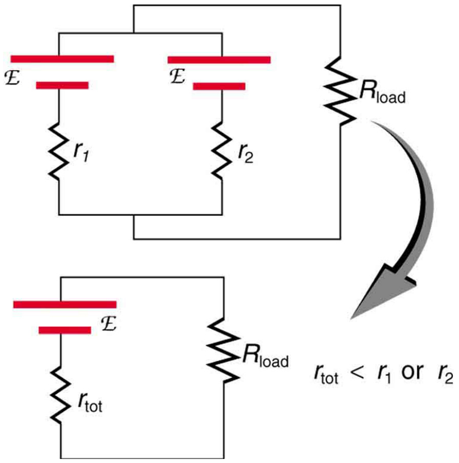
Note:

Take-Home Experiment: Flashlight Batteries

Find a flashlight that uses several batteries and find new and old batteries. Based on the discussions in this module, predict the brightness of the flashlight when different combinations of batteries are used. Do your predictions match what you observe? Now place new batteries in the flashlight and leave the flashlight switched on for several hours. Is the flashlight still quite bright? Do the same with the old batteries. Is the flashlight as bright when left on for the same length of time with old and new batteries? What does this say for the case when you are limited in the number of available new batteries?

[link] shows two voltage sources with identical emfs in parallel and connected to a load resistance. In this simple case, the total emf is the same as the individual emfs. But the total internal resistance is reduced, since the internal resistances are in parallel. The parallel connection thus can produce a larger current.

Here, $I = \frac{\text{emf}}{(r_{\text{tot}} + R_{\text{load}})}$ flows through the load, and r_{tot} is less than those of the individual batteries. For example, some diesel-powered cars use two 12-V batteries in parallel; they produce a total emf of 12 V but can deliver the larger current needed to start a diesel engine.



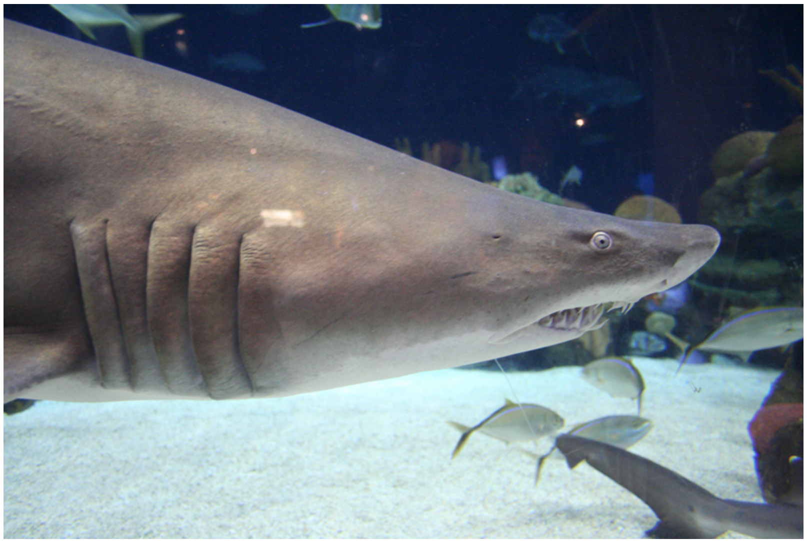
Two voltage sources with identical emfs (each labeled by script E) connected in parallel produce the same emf but have a smaller total internal resistance than the individual sources. Parallel combinations are often used to deliver more current. Here $I = \frac{\text{emf}}{(r_{\text{tot}} + R_{\text{load}})}$ flows through the load.

Animals as Electrical Detectors

A number of animals both produce and detect electrical signals. Fish, sharks, platypuses, and echidnas (spiny anteaters) all detect electric fields generated by nerve activity in prey. Electric eels produce their own emf through biological cells (electric organs) called electroplaques, which are arranged in both series and parallel as a set of batteries.

Electroplaques are flat, disk-like cells; those of the electric eel have a voltage of 0.15 V across each one. These cells are usually located toward the head or tail of the animal, although in the case of the electric eel, they are found along the entire body. The electroplaques in the South American eel are arranged in 140 rows, with each row stretching horizontally along the body and containing 5,000 electroplaques. This can yield an emf of approximately 600 V, and a current of 1 A—deadly.

The mechanism for detection of external electric fields is similar to that for producing nerve signals in the cell through depolarization and repolarization—the movement of ions across the cell membrane. Within the fish, weak electric fields in the water produce a current in a gel-filled canal that runs from the skin to sensing cells, producing a nerve signal. The Australian platypus, one of the very few mammals that lay eggs, can detect fields of 30 $\frac{\text{mV}}{\text{m}}$, while sharks have been found to be able to sense a field in their snouts as small as 100 $\frac{\text{mV}}{\text{m}}$ ([link]). Electric eels use their own electric fields produced by the electroplaques to stun their prey or enemies.



Sand tiger sharks (*Carcharias taurus*), like this one at the Minnesota Zoo, use electroreceptors in their snouts to locate prey. (credit: Jim Winstead, Flickr)

Solar Cell Arrays

Another example dealing with multiple voltage sources is that of combinations of solar cells—wired in both series and parallel combinations to yield a desired voltage and current. Photovoltaic generation (PV), the conversion of sunlight directly into electricity, is based upon the photoelectric effect, in which photons hitting the surface of a solar cell create an electric current in the cell.

Most solar cells are made from pure silicon—either as single-crystal silicon, or as a thin film of silicon deposited upon a glass or metal backing. Most single cells have a voltage output of about 0.5 V, while the current output is a function of the amount of sunlight upon the cell (the incident solar radiation—the insolation). Under bright noon sunlight, a current of about $100\ \text{mA}/\text{cm}^2$ of cell surface area is produced by typical single-crystal cells.

Individual solar cells are connected electrically in modules to meet electrical-energy needs. They can be wired together in series or in parallel—connected like the batteries discussed earlier. A solar-cell array or module usually consists of between 36 and 72 cells, with a power output of 50 W to 140 W.

The output of the solar cells is direct current. For most uses in a home, AC is required, so a device called an inverter must be used to convert the DC to AC. Any extra output can then be passed on to the outside electrical grid for sale to the utility.

Note:
Take-Home Experiment: Virtual Solar Cells
One can assemble a “virtual” solar cell array by using playing cards, or business or index cards, to represent a solar cell. Combinations of these cards in series and/or parallel can model the required array output. Assume each card has an output of 0.5 V and a current (under bright light) of 2 A. Using your cards, how would you arrange them to produce an output of 6 A at 3 V (18 W)?
Suppose you were told that you needed only 18 W (but no required voltage). Would you need more cards to make this arrangement?

Section Summary

- All voltage sources have two fundamental parts—a source of electrical energy that has a characteristic electromotive force (emf), and an internal resistance r.
- The emf is the potential difference of a source when no current is flowing.
- The numerical value of the emf depends on the source of potential difference.
- The internal resistance r of a voltage source affects the output voltage when a current flows.
- The voltage output of a device is called its terminal voltage V and is given by $V=\text{emf}-\text{Ir}$, where I is the electric current and is positive when flowing away from the positive terminal of the voltage source.
- When multiple voltage sources are in series, their internal resistances add and their emfs add algebraically.
- Solar cells can be wired in series or parallel to provide increased voltage or current, respectively.

Conceptual Questions

Exercise:

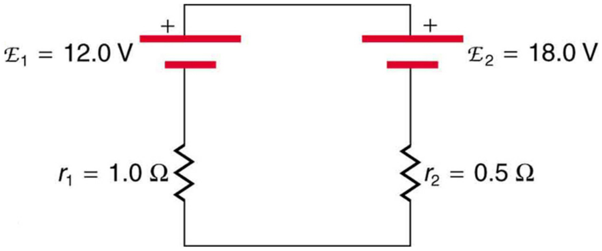
Problem:

Is every emf a potential difference? Is every potential difference an emf? Explain.

Exercise:

Problem:

Explain which battery is doing the charging and which is being charged in [link].



$\mathcal{E}_1 = 12.0\text{ V}$
$\mathcal{E}_2 = 18.0\text{ V}$
$r_1 = 1.0\ \Omega$
$r_2 = 0.5\ \Omega$
+
+

Exercise:

Problem:

Given a battery, an assortment of resistors, and a variety of voltage and current measuring devices, describe how you would determine the internal resistance of the battery.

Exercise:

Problem:

Two different 12-V automobile batteries on a store shelf are rated at 600 and 850 “cold cranking amps.” Which has the smallest internal resistance?

Exercise:

Problem:

What are the advantages and disadvantages of connecting batteries in series? In parallel?

Exercise:

Problem:

Semitractor trucks use four large 12-V batteries. The starter system requires 24 V, while normal operation of the truck’s other electrical components utilizes 12 V. How could the four batteries be connected to produce 24 V? To produce 12 V? Why is 24 V better than 12 V for starting the truck’s engine (a very heavy load)?

Problem Exercises

Exercise:

Problem:

Standard automobile batteries have six lead-acid cells in series, creating a total emf of 12.0 V. What is the emf of an individual lead-acid cell?

Solution:

2.00 V

Exercise:

Problem:

Carbon-zinc dry cells (sometimes referred to as non-alkaline cells) have an emf of 1.54 V, and they are produced as single cells or in various combinations to form other voltages. (a) How many 1.54-V cells are needed to make the common 9-V battery used in many small electronic devices? (b) What is the actual emf of the approximately 9-V battery? (c) Discuss how internal resistance in the series connection of cells will affect the terminal voltage of this approximately 9-V battery.

Exercise:

Problem:

What is the output voltage of a 3.0000-V lithium cell in a digital wristwatch that draws 0.300 mA, if the cell's internal resistance is $2.00\ \Omega$?

Solution:

2.9994 V

Exercise:

Problem:

(a) What is the terminal voltage of a large 1.54-V carbon-zinc dry cell used in a physics lab to supply 2.00 A to a circuit, if the cell's internal resistance is $0.100\ \Omega$? (b) How much electrical power does the cell produce? (c) What power goes to its load?

Exercise:

Problem:

What is the internal resistance of an automobile battery that has an emf of 12.0 V and a terminal voltage of 15.0 V while a current of 8.00 A is charging it?

Solution:

$0.375\ \Omega$

Exercise:

Problem:

(a) Find the terminal voltage of a 12.0-V motorcycle battery having a 0.600-Ω internal resistance, if it is being charged by a current of 10.0 A. (b) What is the output voltage of the battery charger?

Exercise:

Problem:

A car battery with a 12-V emf and an internal resistance of $0.050\,\Omega$ is being charged with a current of 60 A. Note that in this process the battery is being charged. (a) What is the potential difference across its terminals? (b) At what rate is thermal energy being dissipated in the battery? (c) At what rate is electric energy being converted to chemical energy? (d) What are the answers to (a) and (b) when the battery is used to supply 60 A to the starter motor?

Exercise:

Problem:

The hot resistance of a flashlight bulb is $2.30\,\Omega$, and it is run by a 1.58-V alkaline cell having a 0.100-Ω internal resistance. (a) What current flows? (b) Calculate the power supplied to the bulb using I^2R_{bulb}. (c) Is this power the same as calculated using $\frac{V^2}{R_{\text{bulb}}}$?

Solution:

(a) 0.658 A

(b) 0.997 W

(c) 0.997 W; yes

Exercise:

Problem:

The label on a portable radio recommends the use of rechargeable nickel-cadmium cells (nicads), although they have a 1.25-V emf while alkaline cells have a 1.58-V emf. The radio has a 3.20-Ω resistance. (a) Draw a circuit diagram of the radio and its batteries. Now, calculate the power delivered to the radio. (b) When using Nicad cells each having an internal resistance of $0.0400\ \Omega$. (c) When using alkaline cells each having an internal resistance of $0.200\ \Omega$. (d) Does this difference seem significant, considering that the radio's effective resistance is lowered when its volume is turned up?

Exercise:

Problem:

An automobile starter motor has an equivalent resistance of $0.0500\,\Omega$ and is supplied by a 12.0-V battery with a $0.0100\text{-}\Omega$ internal resistance. (a) What is the current to the motor? (b) What voltage is applied to it? (c) What power is supplied to the motor? (d) Repeat these calculations for when the battery connections are corroded and add $0.0900\,\Omega$ to the circuit. (Significant problems are caused by even small amounts of unwanted resistance in low-voltage, high-current applications.)

Solution:

(a) 200 A

(b) 10.0 V

(c) 2.00 kW

(d) $0.1000\ \Omega;\ 80.0\ \text{A},\ 4.0\ \text{V},\ 320\ \text{W}$

Exercise:

Problem:

A child’s electronic toy is supplied by three 1.58-V alkaline cells having internal resistances of $0.0200\,\Omega$ in series with a 1.53-V carbon-zinc dry cell having a $0.100\text{-}\Omega$ internal resistance. The load resistance is $10.0\,\Omega$. (a) Draw a circuit diagram of the toy and its batteries. (b) What current flows? (c) How much power is supplied to the load? (d) What is the internal resistance of the dry cell if it goes bad, resulting in only 0.500 W being supplied to the load?

Exercise:

Problem:

(a) What is the internal resistance of a voltage source if its terminal voltage drops by 2.00 V when the current supplied increases by 5.00 A? (b) Can the emf of the voltage source be found with the information supplied?

Solution:

(a) $0.400\ \Omega$

(b) No, there is only one independent equation, so only r can be found.

Exercise:

Problem:

A person with body resistance between his hands of $10.0\ \text{k}\Omega$ accidentally grasps the terminals of a 20.0-kV power supply. (Do NOT do this!) (a) Draw a circuit diagram to represent the situation. (b) If the internal resistance of the power supply is $2000\ \Omega$, what is the current through his body? (c) What is the power dissipated in his body? (d) If the power supply is to be made safe by increasing its internal resistance, what should the internal resistance be for the maximum current in this situation to be 1.00 mA or less? (e) Will this modification compromise the effectiveness of the power supply for driving low-resistance devices? Explain your reasoning.

Exercise:

Problem:

Electric fish generate current with biological cells called electroplaques, which are physiological emf devices. The electroplaques in the South American eel are arranged in 140 rows, each row stretching horizontally along the body and each containing 5000 electroplaques. Each electroplaque has an emf of 0.15 V and internal resistance of $0.25\ \Omega$. If the water surrounding the fish has resistance of $800\ \Omega$, how much current can the eel produce in water from near its head to near its tail?

Exercise:

Problem: **Integrated Concepts**

A 12.0-V emf automobile battery has a terminal voltage of 16.0 V when being charged by a current of 10.0 A. (a) What is the battery’s internal resistance? (b) What power is dissipated inside the battery? (c) At what rate (in $^{\circ}\text{C/min}$) will its temperature increase if its mass is 20.0 kg and it has a specific heat of $0.300\ \text{kcal/kg}\cdot{}^{\circ}\text{C}$, assuming no heat escapes?

Exercise:

Problem: **Unreasonable Results**

A 1.58-V alkaline cell with a $0.200\text{-}\Omega$ internal resistance is supplying 8.50 A to a load. (a) What is its terminal voltage? (b) What is the value of the load resistance? (c) What is unreasonable about these results? (d) Which assumptions are unreasonable or inconsistent?

Solution:

(a) –0.120 V

(b) $-1.41\times10^{-2}\ \Omega$

(c) Negative terminal voltage; negative load resistance.

(d) The assumption that such a cell could provide 8.50 A is inconsistent with its internal resistance.

Exercise:

Problem: **Unreasonable Results**

(a) What is the internal resistance of a 1.54-V dry cell that supplies 1.00 W of power to a $15.0\text{-}\Omega$ bulb? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Glossary

electromotive force (emf)
the potential difference of a source of electricity when no current is flowing; measured in volts

internal resistance
the amount of resistance within the voltage source

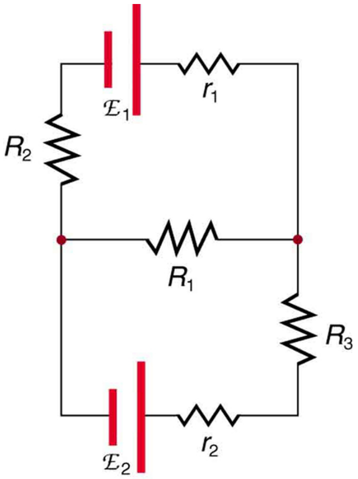
potential difference
the difference in electric potential between two points in an electric circuit, measured in volts

terminal voltage
the voltage measured across the terminals of a source of potential difference

Kirchhoff’s Rules

- Analyze a complex circuit using Kirchhoff’s rules, using the conventions for determining the correct signs of various terms.

Many complex circuits, such as the one in [link], cannot be analyzed with the series-parallel techniques developed in Resistors in Series and Parallel and Electromotive Force: Terminal Voltage. There are, however, two circuit analysis rules that can be used to analyze any circuit, simple or complex. These rules are special cases of the laws of conservation of charge and conservation of energy. The rules are known as **Kirchhoff’s rules**, after their inventor Gustav Kirchhoff (1824–1887).



This circuit cannot be reduced to a combination of series and parallel connections. Kirchhoff’s rules, special applications of the laws of conservation of charge and energy, can be used to analyze it. (Note: The script E in the figure represents electromotive force, emf.)

Note:
Kirchhoff’s Rules

- Kirchhoff’s first rule—the junction rule. The sum of all currents entering a junction must equal the sum of all currents leaving the junction.
- Kirchhoff’s second rule—the loop rule. The algebraic sum of changes in potential around any closed circuit path (loop) must be zero.

Explanations of the two rules will now be given, followed by problem-solving hints for applying Kirchhoff’s rules, and a worked example that uses them.

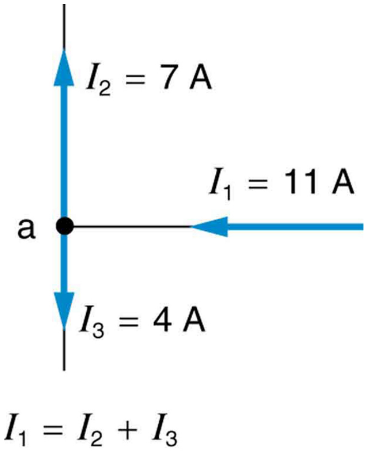
Kirchhoff’s First Rule

Kirchhoff’s first rule (the **junction rule**) is an application of the conservation of charge to a junction; it is illustrated in [link]. Current is the flow of charge, and charge is conserved; thus, whatever charge flows into the junction must flow out. Kirchhoff’s first rule requires that $I_1 = I_2 + I_3$ (see figure). Equations like this can and will be used to analyze circuits and to solve circuit problems.

Note:

Making Connections: Conservation Laws

Kirchhoff’s rules for circuit analysis are applications of **conservation laws** to circuits. The first rule is the application of conservation of charge, while the second rule is the application of conservation of energy. Conservation laws, even used in a specific application, such as circuit analysis, are so basic as to form the foundation of that application.

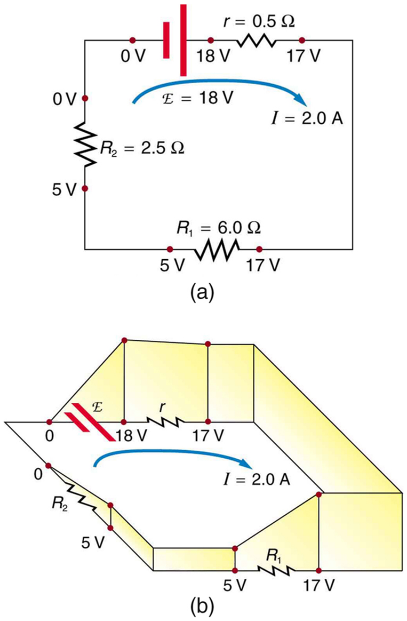


The junction rule. The diagram shows an example of Kirchhoff’s first rule where the sum of the currents into a junction equals the sum of the currents out of a junction. In this case, the current going into the junction splits and comes out as two currents, so that $I_1 = I_2 + I_3$. Here I_1 must be 11 A, since I_2 is 7 A and I_3 is 4 A.

Kirchhoff’s Second Rule

Kirchhoff’s second rule (the **loop rule**) is an application of conservation of energy. The loop rule is stated in terms of potential, V, rather than potential energy, but the two are related since $\mathrm{PE_{elec}} = \mathrm{qV}$. Recall that **emf** is the potential difference of a source when no current is flowing. In a closed loop, whatever energy is supplied by emf must be transferred into other forms by devices in the loop, since there are no other ways in which energy can be transferred into or out of the circuit. [link] illustrates the changes in potential in a simple series circuit loop.

Kirchhoff’s second rule requires $\text{emf} - \mathrm{Ir} - \mathrm{IR_1} - \mathrm{IR_2} = 0$. Rearranged, this is $\text{emf} = \mathrm{Ir} + IR_1 + IR_2$, which means the emf equals the sum of the IR (voltage) drops in the loop.


r = 0.5 Ω
0 V
18 V
17 V
ℰ = 18 V
I = 2.0 A
0 V
R₂ = 2.5 Ω
5 V
R₁ = 6.0 Ω
5 V
17 V
(a)
ℰ
r
0
18 V
17 V
I = 2.0 A
0
R₂
5 V
R₁
5 V
17 V
(b)

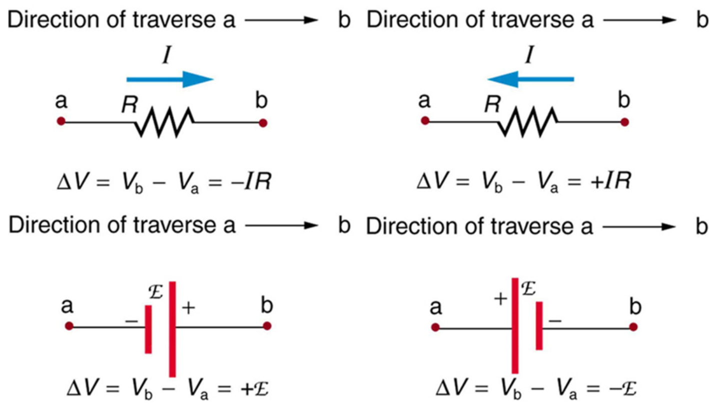
The loop rule. An example of Kirchhoff’s second rule where the sum of the changes in potential around a closed loop must be zero. (a) In this standard schematic of a simple series circuit, the emf supplies 18 V, which is reduced to zero by the resistances, with 1 V across the internal resistance, and 12 V and 5 V across the two load resistances, for a total of 18 V. (b) This perspective view represents the potential as something like a roller coaster, where charge is raised in potential by the emf and lowered by the resistances. (Note that the script E stands for emf.)

Applying Kirchhoff’s Rules

By applying Kirchhoff’s rules, we generate equations that allow us to find the unknowns in circuits. The unknowns may be currents, emfs, or resistances. Each time a rule is applied, an equation is produced. If there are as many independent equations as unknowns, then the problem can be solved. There are two decisions you must make when applying Kirchhoff’s rules. These decisions determine the signs of various quantities in the equations you obtain from applying the rules.

1. When applying Kirchhoff’s first rule, the junction rule, you must label the current in each branch and decide in what direction it is going. For example, in [link], [link], and [link], currents are labeled I_1, I_2, I_3, and I, and arrows indicate their directions. There is no risk here, for if you choose the wrong direction, the current will be of the correct magnitude but negative.
2. When applying Kirchhoff’s second rule, the loop rule, you must identify a closed loop and decide in which direction to go around it, clockwise or counterclockwise. For example, in [link] the loop was traversed in the same direction as the current (clockwise). Again, there is no risk; going around the circuit in the opposite direction reverses the sign of every term in the equation, which is like multiplying both sides of the equation by –1.

[link] and the following points will help you get the plus or minus signs right when applying the loop rule. Note that the resistors and emfs are traversed by going from a to b. In many circuits, it will be necessary to construct more than one loop. In traversing each loop, one needs to be consistent for the sign of the change in potential. (See [link].)


Direction of traverse a → b
I
a
R
b
$\Delta V = V_b - V_a = -IR$
Direction of traverse a → b
I
a
R
b
$\Delta V = V_b - V_a = +IR$
Direction of traverse a → b
a
$\mathcal{E}$
–
+
b
$\Delta V = V_b - V_a = +\mathcal{E}$
Direction of traverse a → b
a
+
$\mathcal{E}$
–
b
$\Delta V = V_b - V_a = -\mathcal{E}$

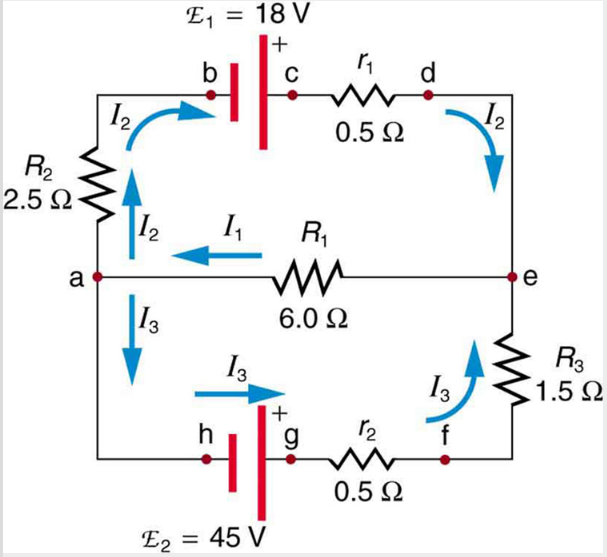
Each of these resistors and voltage sources is traversed from a to b. The potential changes are shown beneath each element and are explained in the text. (Note that the script E stands for emf.)

- When a resistor is traversed in the same direction as the current, the change in potential is $-IR$. (See [link].)
- When a resistor is traversed in the direction opposite to the current, the change in potential is $+IR$. (See [link].)
- When an emf is traversed from – to + (the same direction it moves positive charge), the change in potential is +emf. (See [link].)
- When an emf is traversed from + to – (opposite to the direction it moves positive charge), the change in potential is $-$emf. (See [link].)

Example:

Calculating Current: Using Kirchhoff’s Rules

Find the currents flowing in the circuit in [link].



This circuit is similar to that in [link], but the resistances and emfs are specified. (Each emf is denoted by script E.) The currents in each branch are labeled and assumed to move in the directions shown. This example uses Kirchhoff’s rules to find the currents.

Strategy

This circuit is sufficiently complex that the currents cannot be found using Ohm’s law and the series-parallel techniques—it is necessary to use Kirchhoff’s rules. Currents have been labeled I_1, I_2, and I_3 in the figure and assumptions have been made about their directions. Locations on the diagram have been labeled with letters a through h. In the solution we will apply the junction and loop rules, seeking three independent equations to allow us to solve for the three unknown currents.

Solution

We begin by applying Kirchhoff's first or junction rule at point a. This gives

Equation:

$$I_1 = I_2 + I_3,$$

since I_1 flows into the junction, while I_2 and I_3 flow out. Applying the junction rule at e produces exactly the same equation, so that no new information is obtained. This is a single equation with three unknowns—three independent equations are needed, and so the loop rule must be applied.

Now we consider the loop abcdea. Going from a to b, we traverse R_2 in the same (assumed) direction of the current I_2, and so the change in potential is $-I_2R_2$. Then going from b to c, we go from – to +, so that the change in potential is $+\text{emf}_1$. Traversing the internal resistance r_1 from c to d gives $-I_2r_1$. Completing the loop by going from d to a again traverses a resistor in the same direction as its current, giving a change in potential of $-I_1R_1$.

The loop rule states that the changes in potential sum to zero. Thus,

Equation:

$$-I_2R_2 + \text{emf}_1 - I_2r_1 - I_1R_1 = -I_2(R_2 + r_1) + \text{emf}_1 - I_1R_1 = 0.$$

Substituting values from the circuit diagram for the resistances and emf, and canceling the ampere unit gives

Equation:

$$-3I_2 + 18 - 6I_1 = 0.$$

Now applying the loop rule to aefgha (we could have chosen abcdefgha as well) similarly gives

Equation:

$$+ I_1R_1 + I_3R_3 + I_3r_2 - \text{emf}_2 = +I_1R_1 + I_3(R_3 + r_2) - \text{emf}_2 = 0.$$

Note that the signs are reversed compared with the other loop, because elements are traversed in the opposite direction. With values entered, this becomes

Equation:

$$+ 6I_1 + 2I_3 - 45 = 0.$$

These three equations are sufficient to solve for the three unknown currents. First, solve the second equation for I_2:

Equation:

$$I_2 = 6 - 2I_1.$$

Now solve the third equation for I_3:

Equation:

$$I_3 = 22.5 - 3I_1.$$

Substituting these two new equations into the first one allows us to find a value for I_1:

Equation:

$$I_1 = I_2 + I_3 = (6 - 2I_1) + (22.5 - 3I_1) = 28.5 - 5I_1.$$

Combining terms gives

Equation:

$$6I_1 = 28.5, \text{ and}$$

Equation:

$$I_1 = 4.75 \text{ A}.$$

Substituting this value for I_1 back into the fourth equation gives

Equation:

$$I_2 = 6 - 2I_1 = 6 - 9.50$$

Equation:

$$I_2 = -3.50 \text{ A}.$$

The minus sign means I_2 flows in the direction opposite to that assumed in [link].

Finally, substituting the value for I_1 into the fifth equation gives

Equation:

$$I_3 = 22.5 - 3I_1 = 22.5 - 14.25$$

Equation:

$$I_3 = 8.25 \text{ A}.$$

Discussion

Just as a check, we note that indeed $I_1 = I_2 + I_3$. The results could also have been checked by entering all of the values into the equation for the abcdefgha loop.

Note:

Problem-Solving Strategies for Kirchhoff’s Rules

1. Make certain there is a clear circuit diagram on which you can label all known and unknown resistances, emfs, and currents. If a current is unknown, you must assign it a direction. This is necessary for determining the signs of potential changes. If you assign the direction incorrectly, the current will be found to have a negative value—no harm done.
2. Apply the junction rule to any junction in the circuit. Each time the junction rule is applied, you should get an equation with a current that does not appear in a previous application—if not, then the equation is redundant.
3. Apply the loop rule to as many loops as needed to solve for the unknowns in the problem. (There must be as many independent equations as unknowns.) To apply the loop rule, you must choose a direction to go around the loop. Then carefully and consistently determine the signs of the potential changes for each element using the four bulleted points discussed above in conjunction with [link].
4. Solve the simultaneous equations for the unknowns. This may involve many algebraic steps, requiring careful checking and rechecking.
5. Check to see whether the answers are reasonable and consistent. The numbers should be of the correct order of magnitude, neither

exceedingly large nor vanishingly small. The signs should be reasonable—for example, no resistance should be negative. Check to see that the values obtained satisfy the various equations obtained from applying the rules. The currents should satisfy the junction rule, for example.

The material in this section is correct in theory. We should be able to verify it by making measurements of current and voltage. In fact, some of the devices used to make such measurements are straightforward applications of the principles covered so far and are explored in the next modules. As we shall see, a very basic, even profound, fact results—making a measurement alters the quantity being measured.

Exercise:

Check Your Understanding

Problem:

Can Kirchhoff’s rules be applied to simple series and parallel circuits or are they restricted for use in more complicated circuits that are not combinations of series and parallel?

Solution:

Kirchhoff's rules can be applied to any circuit since they are applications to circuits of two conservation laws. Conservation laws are the most broadly applicable principles in physics. It is usually mathematically simpler to use the rules for series and parallel in simpler circuits so we emphasize Kirchhoff’s rules for use in more complicated situations. But the rules for series and parallel can be derived from Kirchhoff’s rules. Moreover, Kirchhoff’s rules can be expanded to devices other than resistors and emfs, such as capacitors, and are one of the basic analysis devices in circuit analysis.

Section Summary

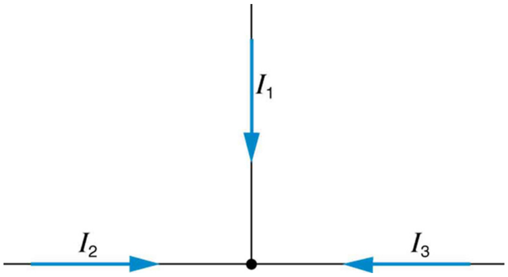
- Kirchhoff’s rules can be used to analyze any circuit, simple or complex.
- Kirchhoff’s first rule—the junction rule: The sum of all currents entering a junction must equal the sum of all currents leaving the junction.
- Kirchhoff’s second rule—the loop rule: The algebraic sum of changes in potential around any closed circuit path (loop) must be zero.
- The two rules are based, respectively, on the laws of conservation of charge and energy.
- When calculating potential and current using Kirchhoff’s rules, a set of conventions must be followed for determining the correct signs of various terms.
- The simpler series and parallel rules are special cases of Kirchhoff’s rules.

Conceptual Questions

Exercise:

Problem:

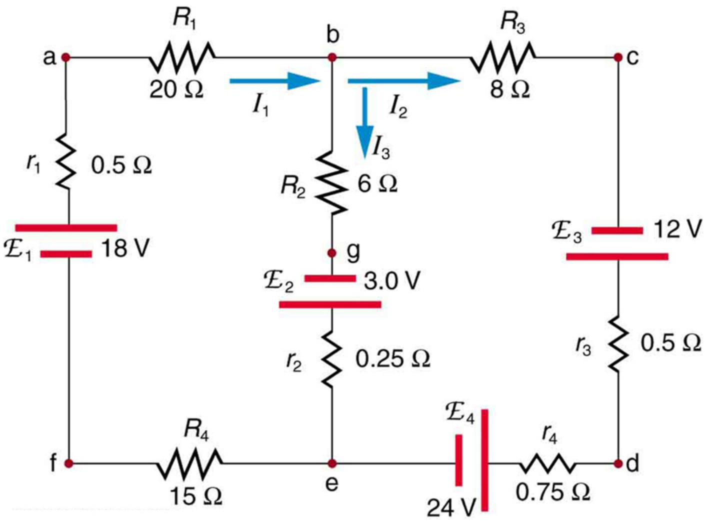
Can all of the currents going into the junction in [link] be positive? Explain.



Exercise:

Problem:

Apply the junction rule to junction b in [link]. Is any new information gained by applying the junction rule at e? (In the figure, each emf is represented by script E.)



R_1
b
R_3
a
c
20 Ω
I_1
I_2
8 Ω
I_3
r_1
0.5 Ω
R_2
6 Ω
$\mathcal{E}_3$
12 V
$\mathcal{E}_1$
18 V
g
$\mathcal{E}_2$
3.0 V
r_3
0.5 Ω
r_2
0.25 Ω
$\mathcal{E}_4$
R_4
r_4
f
d
e
15 Ω
24 V
0.75 Ω

Exercise:

Problem:

(a) What is the potential difference going from point a to point b in [link]? (b) What is the potential difference going from c to b? (c) From e to g? (d) From e to d?

Exercise:

Problem: Apply the loop rule to loop afedcba in [link].

Exercise:

Problem: Apply the loop rule to loops abgefa and cbgedc in [link].

Problem Exercises

Exercise:

Problem: Apply the loop rule to loop abcdefgha in [link].

Solution:
Equation:

$$-I_2R_2+\text{emf}_1-\text{I}_2r_1+\text{I}_3R_3+\text{I}_3r_2-\text{emf}_2=0$$

Exercise:

Problem: Apply the loop rule to loop aedcba in [link].

Exercise:

Problem:

Verify the second equation in [link] by substituting the values found for the currents I_1 and I_2.

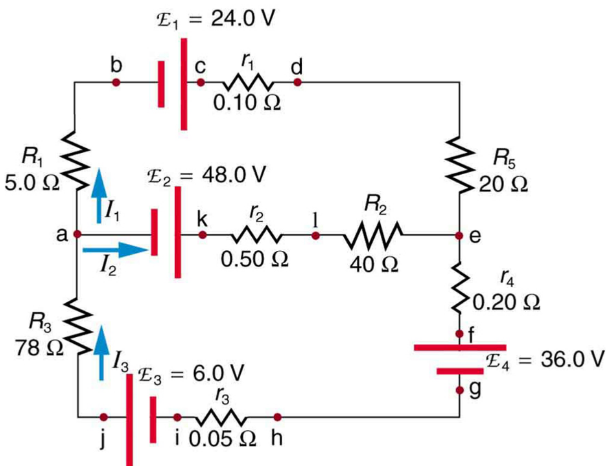
Exercise:

Problem:

Verify the third equation in [link] by substituting the values found for the currents I_1 and I_3.

Exercise:

Problem: Apply the junction rule at point a in [link].


$\mathcal{E}_1 = 24.0\ \text{V}$
b
c
r_1
d
$0.10\ \Omega$
R_1
$5.0\ \Omega$
$\mathcal{E}_2 = 48.0\ \text{V}$
R_5
$20\ \Omega$
I_1
R_2
k
r_2
l
a
e
$0.50\ \Omega$
$40\ \Omega$
I_2
r_4
$0.20\ \Omega$
R_3
$78\ \Omega$
f
I_3
$\mathcal{E}_3 = 6.0\ \text{V}$
$\mathcal{E}_4 = 36.0\ \text{V}$
r_3
g
j
i
$0.05\ \Omega$
h

Solution:
Equation:

$$I_3 = I_1 + I_2$$

Exercise:

Problem: Apply the loop rule to loop abcdefghija in [link].

Exercise:

Problem: Apply the loop rule to loop akledcba in [link].

Solution:
Equation:

$$\text{emf}_2 - I_2 r_2 - I_2 R_2 + I_1 R_5 + I_1 r_1 - \text{emf}_1 + I_1 R_1 = 0$$

Exercise:

Problem:

Find the currents flowing in the circuit in [link]. Explicitly show how you follow the steps in the Problem-Solving Strategies for Series and Parallel Resistors.

Exercise:

Problem:

Solve [link], but use loop abcdefgha instead of loop akledcba. Explicitly show how you follow the steps in the Problem-Solving Strategies for Series and Parallel Resistors.

Solution:

(a) $\mathrm{I}_1 = 4.75\text{ A}$

(b) $\mathrm{I}_2 = -3.5\text{ A}$

(c) $\mathrm{I}_3 = 8.25\text{ A}$

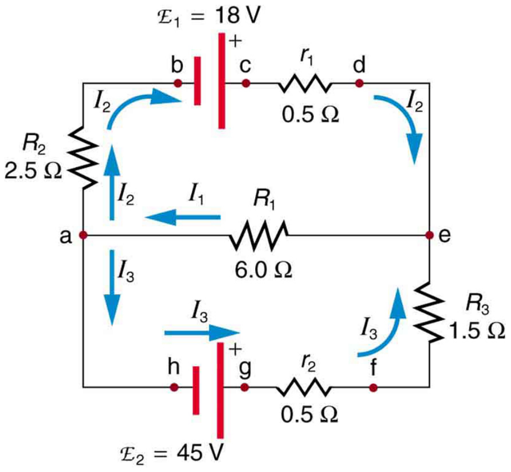
Exercise:

Problem: Find the currents flowing in the circuit in [link].

Exercise:

Problem: Unreasonable Results

Consider the circuit in [link], and suppose that the emfs are unknown and the currents are given to be $I_1 = 5.00\text{ A}$, $I_2 = 3.0\text{ A}$, and $I_3 = -2.00\text{ A}$. (a) Could you find the emfs? (b) What is wrong with the assumptions?


$\mathcal{E}_1 = 18$ V
+
b
c
r_1
d
I_2
0.5 Ω
I_2
R_2
2.5 Ω
I_2
I_1
R_1
a
e
6.0 Ω
I_3
I_3
R_3
1.5 Ω
I_3
+
h
g
r_2
f
0.5 Ω
$\mathcal{E}_2 = 45$ V

Solution:

(a) No, you would get inconsistent equations to solve.

(b) $I_1 \neq I_2 + I_3$. The assumed currents violate the junction rule.

Glossary

Kirchhoff’s rules

a set of two rules, based on conservation of charge and energy, governing current and changes in potential in an electric circuit

junction rule

Kirchhoff’s first rule, which applies the conservation of charge to a junction; current is the flow of charge; thus, whatever charge flows into the junction must flow out; the rule can be stated $I_1 = I_2 + I_3$

loop rule

Kirchhoff’s second rule, which states that in a closed loop, whatever energy is supplied by emf must be transferred into other forms by devices in the loop, since there are no other ways in which energy can be transferred into or out of the circuit. Thus, the emf equals the sum of the IR (voltage) drops in the loop and can be stated: emf $= Ir + IR_1 + IR_2$

conservation laws

require that energy and charge be conserved in a system

DC Voltmeters and Ammeters

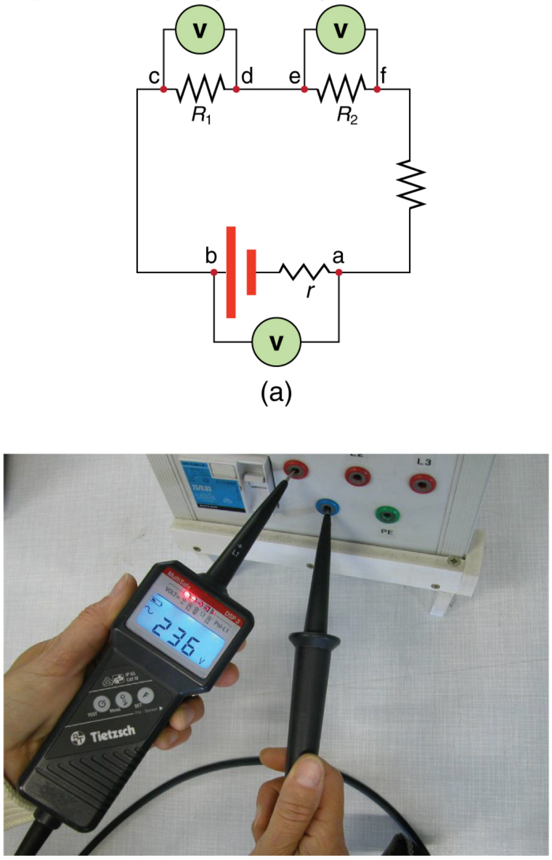
- Explain why a voltmeter must be connected in parallel with the circuit.
- Draw a diagram showing an ammeter correctly connected in a circuit.
- Describe how a galvanometer can be used as either a voltmeter or an ammeter.
- Find the resistance that must be placed in series with a galvanometer to allow it to be used as a voltmeter with a given reading.
- Explain why measuring the voltage or current in a circuit can never be exact.

Voltmeters measure voltage, whereas **ammeters** measure current. Some of the meters in automobile dashboards, digital cameras, cell phones, and tuner-amplifiers are voltmeters or ammeters. (See [link].) The internal construction of the simplest of these meters and how they are connected to the system they monitor give further insight into applications of series and parallel connections.

The fuel and temperature gauges (far right and far left, respectively) in this 1996 Volkswagen are voltmeters that register the voltage output of "sender" units, which are hopefully proportional to the amount of gasoline in the tank and the engine temperature. (credit: Christian Giersing)

Voltmeters are connected in parallel with whatever device’s voltage is to be measured. A parallel connection is used because objects in parallel experience the same potential difference. (See [link], where the voltmeter is represented by the symbol V.)

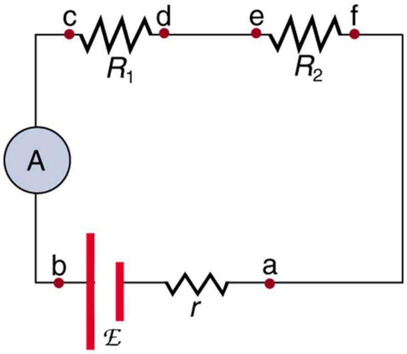
Ammeters are connected in series with whatever device’s current is to be measured. A series connection is used because objects in series have the same current passing through them. (See [link], where the ammeter is represented by the symbol A.)



(a)

(b)

(a) To measure potential differences in this series circuit, the voltmeter (V) is placed in parallel with the voltage source or either of the resistors. Note that terminal voltage is measured between points a and b. It is not possible to connect the voltmeter directly across the emf without including its internal resistance, r. (b) A digital voltmeter in use. (credit: Messtechniker, Wikimedia Commons)



An ammeter (A) is placed in series to measure current. All of the current in this circuit flows through the meter. The ammeter would have the same reading if located between points d and e or between points f and a as it does in the position shown. (Note that the script capital E stands for emf, and r stands for the internal resistance of the source of potential difference.)

Analog Meters: Galvanometers

Analog meters have a needle that swivels to point at numbers on a scale, as opposed to **digital meters**, which have numerical readouts similar to a hand-held calculator. The heart of most analog meters is a device called a **galvanometer**, denoted by G. Current flow through a galvanometer, I_{G}, produces a proportional needle deflection. (This deflection is due to the force of a magnetic field upon a current-carrying wire.)

The two crucial characteristics of a given galvanometer are its resistance and current sensitivity. **Current sensitivity** is the current that gives a **full-scale deflection** of the galvanometer's needle, the maximum current that the instrument can measure. For example, a galvanometer with a current sensitivity of $50\ \mu\text{A}$ has a maximum deflection of its needle when $50\ \mu\text{A}$ flows through it, reads half-scale when $25\ \mu\text{A}$ flows through it, and so on.

If such a galvanometer has a $25\text{-}\Omega$ resistance, then a voltage of only $V=\text{IR}=(50\ \mu\text{A})(25\ \Omega)=1.25\ \text{mV}$ produces a full-scale reading. By connecting resistors to this galvanometer in different ways, you can use it as either a voltmeter or ammeter that can measure a broad range of voltages or currents.

Galvanometer as Voltmeter

[link] shows how a galvanometer can be used as a voltmeter by connecting it in series with a large resistance, R. The value of the resistance R is

determined by the maximum voltage to be measured. Suppose you want 10 V to produce a full-scale deflection of a voltmeter containing a $25\text{-}\Omega$ galvanometer with a $50\text{-}\mu\text{A}$ sensitivity. Then 10 V applied to the meter must produce a current of $50\ \mu\text{A}$. The total resistance must be

Equation:

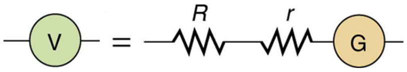
$$R_{\text{tot}} = R + r = \frac{V}{I} = \frac{10\text{ V}}{50\ \mu\text{A}} = 200\text{ k }\Omega\text{, or}$$

Equation:

$$R = R_{\text{tot}} - r = 200\text{ k}\Omega - 25\ \Omega \approx 200\text{ k}\Omega.$$

(R is so large that the galvanometer resistance, r, is nearly negligible.) Note that 5 V applied to this voltmeter produces a half-scale deflection by producing a $25\text{-}\mu\text{A}$ current through the meter, and so the voltmeter's reading is proportional to voltage as desired.

This voltmeter would not be useful for voltages less than about half a volt, because the meter deflection would be small and difficult to read accurately. For other voltage ranges, other resistances are placed in series with the galvanometer. Many meters have a choice of scales. That choice involves switching an appropriate resistance into series with the galvanometer.



A large resistance R placed in series with a galvanometer G produces a voltmeter, the full-scale deflection of which depends on the choice of R.

The larger the voltage to be measured, the larger R must be. (Note that r represents the internal resistance of the galvanometer.)

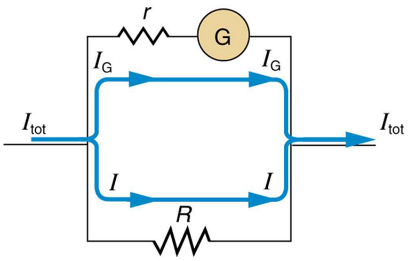
Galvanometer as Ammeter

The same galvanometer can also be made into an ammeter by placing it in parallel with a small resistance R, often called the **shunt resistance**, as shown in [link]. Since the shunt resistance is small, most of the current passes through it, allowing an ammeter to measure currents much greater than those producing a full-scale deflection of the galvanometer.

Suppose, for example, an ammeter is needed that gives a full-scale deflection for 1.0 A, and contains the same $25\text{-}\Omega$ galvanometer with its 50-μA sensitivity. Since R and r are in parallel, the voltage across them is the same.

These IR drops are $\text{IR} = I_\text{G}r$ so that $\text{IR} = \frac{I_\text{G}}{I} = \frac{R}{r}$. Solving for R, and noting that I_G is 50 μA and I is 0.999950 A, we have
Equation:

$$R = r\frac{I_\text{G}}{I} = (25\,\Omega)\frac{50\ \mu\text{A}}{0.999950\ \text{A}} = 1.25\times10^{-3}\,\Omega.$$

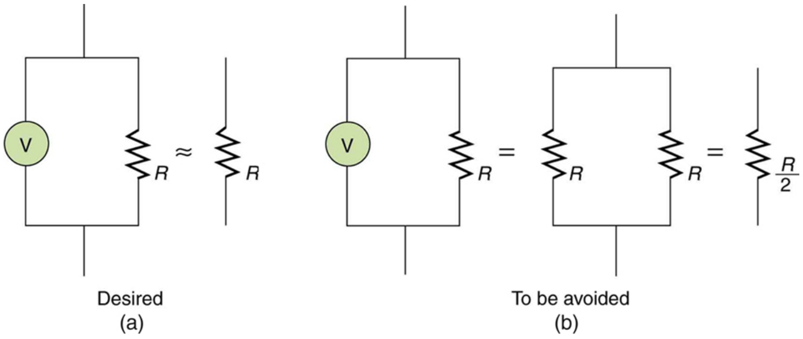


A small shunt resistance R placed in parallel with a galvanometer G produces an ammeter, the full-scale deflection of which depends on the choice of R. The larger the current to be measured, the smaller R must be. Most of the current (I) flowing through the meter is shunted through R to protect the galvanometer. (Note that r represents the internal resistance of the galvanometer.) Ammeters may also have multiple scales for greater flexibility in application. The various scales are achieved by switching various shunt resistances in parallel with the galvanometer—the greater the maximum current to be measured, the smaller the shunt resistance must be.

Taking Measurements Alters the Circuit

When you use a voltmeter or ammeter, you are connecting another resistor to an existing circuit and, thus, altering the circuit. Ideally, voltmeters and ammeters do not appreciably affect the circuit, but it is instructive to examine the circumstances under which they do or do not interfere.

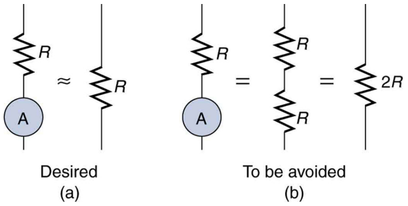
First, consider the voltmeter, which is always placed in parallel with the device being measured. Very little current flows through the voltmeter if its resistance is a few orders of magnitude greater than the device, and so the circuit is not appreciably affected. (See [link](a).) (A large resistance in parallel with a small one has a combined resistance essentially equal to the small one.) If, however, the voltmeter’s resistance is comparable to that of the device being measured, then the two in parallel have a smaller resistance, appreciably affecting the circuit. (See [link](b).) The voltage across the device is not the same as when the voltmeter is out of the circuit.



(a) A voltmeter having a resistance much larger than the device ($R_{\text{Voltmeter}}>>R$) with which it is in parallel produces a parallel resistance essentially the same as the device and does not appreciably affect the circuit being measured. (b) Here the voltmeter has the same resistance as the device ($R_{\text{Voltmeter}} \cong R$), so that the parallel resistance is half of what it is when the voltmeter is not connected. This is an example of a significant alteration of the circuit and is to be avoided.

An ammeter is placed in series in the branch of the circuit being measured, so that its resistance adds to that branch. Normally, the ammeter's resistance is very small compared with the resistances of the devices in the circuit, and so the extra resistance is negligible. (See [link](a).) However, if very small load resistances are involved, or if the ammeter is not as low in resistance as it should be, then the total series resistance is significantly greater, and the current in the branch being measured is reduced. (See [link](b).)

A practical problem can occur if the ammeter is connected incorrectly. If it was put in parallel with the resistor to measure the current in it, you could possibly damage the meter; the low resistance of the ammeter would allow most of the current in the circuit to go through the galvanometer, and this current would be larger since the effective resistance is smaller.



(a) An ammeter normally has such a small resistance that the total series resistance in the branch being measured is not appreciably increased. The circuit is essentially unaltered compared with when the ammeter is absent. (b) Here the ammeter’s resistance is the same as that of the branch, so that the total resistance is doubled and the current is half what it is without the ammeter. This significant alteration of the circuit is to be avoided.

One solution to the problem of voltmeters and ammeters interfering with the circuits being measured is to use galvanometers with greater sensitivity.

This allows construction of voltmeters with greater resistance and ammeters with smaller resistance than when less sensitive galvanometers are used.

There are practical limits to galvanometer sensitivity, but it is possible to get analog meters that make measurements accurate to a few percent. Note that the inaccuracy comes from altering the circuit, not from a fault in the meter.

Note:

Connections: Limits to Knowledge

Making a measurement alters the system being measured in a manner that produces uncertainty in the measurement. For macroscopic systems, such as the circuits discussed in this module, the alteration can usually be made negligibly small, but it cannot be eliminated entirely. For submicroscopic systems, such as atoms, nuclei, and smaller particles, measurement alters the system in a manner that cannot be made arbitrarily small. This actually limits knowledge of the system—even limiting what nature can know about itself. We shall see profound implications of this when the Heisenberg uncertainty principle is discussed in the modules on quantum mechanics.

There is another measurement technique based on drawing no current at all and, hence, not altering the circuit at all. These are called null measurements and are the topic of Null Measurements. Digital meters that employ solid-state electronics and null measurements can attain accuracies of one part in 10^6.

Exercise:

Check Your Understanding

Problem:

Digital meters are able to detect smaller currents than analog meters employing galvanometers. How does this explain their ability to measure voltage and current more accurately than analog meters?

Solution:

Since digital meters require less current than analog meters, they alter the circuit less than analog meters. Their resistance as a voltmeter can be far greater than an analog meter, and their resistance as an ammeter can be far less than an analog meter. Consult [link] and [link] and their discussion in the text.

Note:
PhET Explorations: Circuit Construction Kit (DC Only), Virtual Lab
Stimulate a neuron and monitor what happens. Pause, rewind, and move forward in time in order to observe the ions as they move across the neuron membrane.

Circuit Construction Kit (DC Only), Virtual Lab

Section Summary

- Voltmeters measure voltage, and ammeters measure current.
- A voltmeter is placed in parallel with the voltage source to receive full voltage and must have a large resistance to limit its effect on the circuit.
- An ammeter is placed in series to get the full current flowing through a branch and must have a small resistance to limit its effect on the circuit.

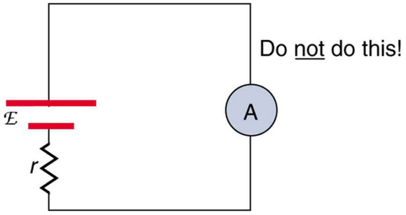
- Both can be based on the combination of a resistor and a galvanometer, a device that gives an analog reading of current.
- Standard voltmeters and ammeters alter the circuit being measured and are thus limited in accuracy.

Conceptual Questions

Exercise:

Problem:

Why should you not connect an ammeter directly across a voltage source as shown in [link]? (Note that script E in the figure stands for emf.)



Exercise:

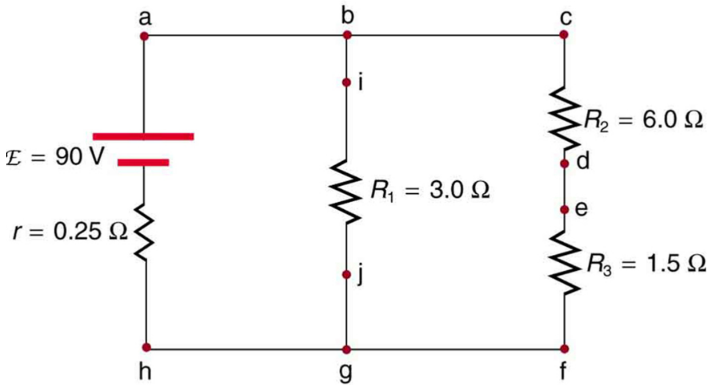
Problem:

Suppose you are using a multimeter (one designed to measure a range of voltages, currents, and resistances) to measure current in a circuit and you inadvertently leave it in a voltmeter mode. What effect will the meter have on the circuit? What would happen if you were measuring voltage but accidentally put the meter in the ammeter mode?

Exercise:

Problem:

Specify the points to which you could connect a voltmeter to measure the following potential differences in [link]: (a) the potential difference of the voltage source; (b) the potential difference across R_1; (c) across R_2; (d) across R_3; (e) across R_2 and R_3. Note that there may be more than one answer to each part.



Exercise:

Problem:

To measure currents in [link], you would replace a wire between two points with an ammeter. Specify the points between which you would place an ammeter to measure the following: (a) the total current; (b) the current flowing through R_1; (c) through R_2; (d) through R_3. Note that there may be more than one answer to each part.

Problem Exercises

Exercise:

Problem:

What is the sensitivity of the galvanometer (that is, what current gives a full-scale deflection) inside a voltmeter that has a $1.00\text{-M}\Omega$ resistance on its 30.0-V scale?

Solution:

$30\ \mu A$

Exercise:

Problem:

What is the sensitivity of the galvanometer (that is, what current gives a full-scale deflection) inside a voltmeter that has a $25.0\text{-k}\Omega$ resistance on its 100-V scale?

Exercise:

Problem:

Find the resistance that must be placed in series with a $25.0\text{-}\Omega$ galvanometer having a $50.0-\mu\text{A}$ sensitivity (the same as the one discussed in the text) to allow it to be used as a voltmeter with a 0.100-V full-scale reading.

Solution:

1.98 kΩ

Exercise:

Problem:

Find the resistance that must be placed in series with a $25.0\text{-}\Omega$ galvanometer having a $50.0\text{-}\mu\text{A}$ sensitivity (the same as the one discussed in the text) to allow it to be used as a voltmeter with a 3000-V full-scale reading. Include a circuit diagram with your solution.

Exercise:

Problem:

Find the resistance that must be placed in parallel with a $25.0\text{-}\Omega$ galvanometer having a $50.0\text{-}\mu\text{A}$ sensitivity (the same as the one discussed in the text) to allow it to be used as an ammeter with a 10.0-A full-scale reading. Include a circuit diagram with your solution.

Solution:
Equation:

$$1.25\times10^{-4}\ \Omega$$

Exercise:

Problem:

Find the resistance that must be placed in parallel with a $25.0\text{-}\Omega$ galvanometer having a $50.0\text{-}\mu\text{A}$ sensitivity (the same as the one discussed in the text) to allow it to be used as an ammeter with a 300-mA full-scale reading.

Exercise:

Problem:

Find the resistance that must be placed in series with a $10.0\text{-}\Omega$ galvanometer having a $100\text{-}\mu\text{A}$ sensitivity to allow it to be used as a voltmeter with: (a) a 300-V full-scale reading, and (b) a 0.300-V full-scale reading.

Solution:

(a) $3.00\ \text{M}\Omega$

(b) $2.99\ \text{k}\Omega$

Exercise:

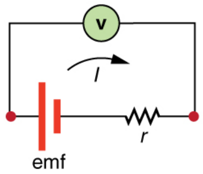
Problem:

Find the resistance that must be placed in parallel with a $10.0\text{-}\Omega$ galvanometer having a $100\text{-}\mu\text{A}$ sensitivity to allow it to be used as an ammeter with: (a) a 20.0-A full-scale reading, and (b) a 100-mA full-scale reading.

Exercise:

Problem:

Suppose you measure the terminal voltage of a 1.585-V alkaline cell having an internal resistance of $0.100\,\Omega$ by placing a $1.00\text{-k}\Omega$ voltmeter across its terminals. (See [link].) (a) What current flows? (b) Find the terminal voltage. (c) To see how close the measured terminal voltage is to the emf, calculate their ratio.



Solution:

(a) 1.58 mA

(b) 1.5848 V (need four digits to see the difference)

(c) 0.99990 (need five digits to see the difference from unity)

Exercise:

Problem:

Suppose you measure the terminal voltage of a 3.200-V lithium cell having an internal resistance of $5.00\,\Omega$ by placing a $1.00\text{-k}\Omega$ voltmeter across its terminals. (a) What current flows? (b) Find the terminal voltage. (c) To see how close the measured terminal voltage is to the emf, calculate their ratio.

Exercise:

Problem:

A certain ammeter has a resistance of $5.00\times10^{-5}\ \Omega$ on its 3.00-A scale and contains a $10.0\text{-}\Omega$ galvanometer. What is the sensitivity of the galvanometer?

Solution:

15.0 µA

Exercise:

Problem:

A 1.00-MΩ voltmeter is placed in parallel with a $75.0\text{-k}\Omega$ resistor in a circuit. (a) Draw a circuit diagram of the connection. (b) What is the resistance of the combination? (c) If the voltage across the combination is kept the same as it was across the $75.0\text{-k}\Omega$ resistor alone, what is the percent increase in current? (d) If the current through the combination is kept the same as it was through the $75.0\text{-k}\Omega$ resistor alone, what is the percentage decrease in voltage? (e) Are the changes found in parts (c) and (d) significant? Discuss.

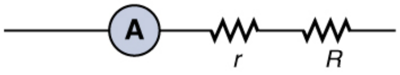
Exercise:

Problem:

A 0.0200-Ω ammeter is placed in series with a $10.00\text{-}\Omega$ resistor in a circuit. (a) Draw a circuit diagram of the connection. (b) Calculate the resistance of the combination. (c) If the voltage is kept the same across the combination as it was through the $10.00\text{-}\Omega$ resistor alone, what is the percent decrease in current? (d) If the current is kept the same through the combination as it was through the $10.00\text{-}\Omega$ resistor alone, what is the percent increase in voltage? (e) Are the changes found in parts (c) and (d) significant? Discuss.

Solution:

(a)



(b) 10.02 Ω

(c) 0.9980, or a 2.0×10^{-1} percent decrease

(d) 1.002, or a 2.0×10^{-1} percent increase

(e) Not significant.

Exercise:

Problem: Unreasonable Results

Suppose you have a $40.0\text{-}\Omega$ galvanometer with a $25.0\text{-}\mu\text{A}$ sensitivity. (a) What resistance would you put in series with it to allow it to be used as a voltmeter that has a full-scale deflection for $0.500\ \text{mV}$? (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem: Unreasonable Results

(a) What resistance would you put in parallel with a $40.0\text{-}\Omega$ galvanometer having a $25.0\text{-}\mu\text{A}$ sensitivity to allow it to be used as an ammeter that has a full-scale deflection for $10.0\text{-}\mu\text{A}$? (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Solution:

(a) $-66.7\ \Omega$

(b) You can't have negative resistance.

(c) It is unreasonable that I_{G} is greater than I_{tot} (see [link]). You cannot achieve a full-scale deflection using a current less than the sensitivity of the galvanometer.

Glossary

voltmeter
 an instrument that measures voltage

ammeter

an instrument that measures current

analog meter
a measuring instrument that gives a readout in the form of a needle movement over a marked gauge

digital meter
a measuring instrument that gives a readout in a digital form

galvanometer
an analog measuring device, denoted by G, that measures current flow using a needle deflection caused by a magnetic field force acting upon a current-carrying wire

current sensitivity
the maximum current that a galvanometer can read

full-scale deflection
the maximum deflection of a galvanometer needle, also known as current sensitivity; a galvanometer with a full-scale deflection of $50\ \mu\text{A}$ has a maximum deflection of its needle when $50\ \mu\text{A}$ flows through it

shunt resistance
a small resistance R placed in parallel with a galvanometer G to produce an ammeter; the larger the current to be measured, the smaller R must be; most of the current flowing through the meter is shunted through R to protect the galvanometer

Null Measurements

- Explain why a null measurement device is more accurate than a standard voltmeter or ammeter.
- Demonstrate how a Wheatstone bridge can be used to accurately calculate the resistance in a circuit.

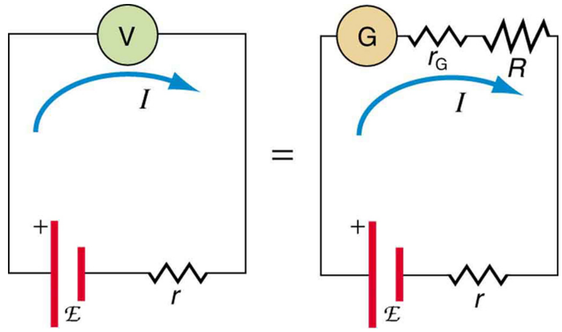
Standard measurements of voltage and current alter the circuit being measured, introducing uncertainties in the measurements. Voltmeters draw some extra current, whereas ammeters reduce current flow. **Null measurements** balance voltages so that there is no current flowing through the measuring device and, therefore, no alteration of the circuit being measured.

Null measurements are generally more accurate but are also more complex than the use of standard voltmeters and ammeters, and they still have limits to their precision. In this module, we shall consider a few specific types of null measurements, because they are common and interesting, and they further illuminate principles of electric circuits.

The Potentiometer

Suppose you wish to measure the emf of a battery. Consider what happens if you connect the battery directly to a standard voltmeter as shown in [link]. (Once we note the problems with this measurement, we will examine a null measurement that improves accuracy.) As discussed before, the actual quantity measured is the terminal voltage V, which is related to the emf of the battery by $V=\text{emf}-\text{Ir}$, where I is the current that flows and r is the internal resistance of the battery.

The emf could be accurately calculated if r were very accurately known, but it is usually not. If the current I could be made zero, then $V=\text{emf}$, and so emf could be directly measured. However, standard voltmeters need a current to operate; thus, another technique is needed.



An analog voltmeter attached to a battery draws a small but nonzero current and measures a terminal voltage that differs from the emf of the battery. (Note that the script capital E symbolizes electromotive force, or emf.) Since the internal resistance of the battery is not known precisely, it is not possible to calculate the emf precisely.

A **potentiometer** is a null measurement device for measuring potentials (voltages). (See [link].) A voltage source is connected to a resistor R, say, a long wire, and passes a constant current through it. There is a steady drop in potential (an IR drop) along the wire, so that a variable potential can be obtained by making contact at varying locations along the wire.

[link](b) shows an unknown $\mathrm{emf}_{\mathrm{x}}$ (represented by script E_{x} in the figure) connected in series with a galvanometer. Note that $\mathrm{emf}_{\mathrm{x}}$ opposes the other voltage source. The location of the contact point (see the arrow on the drawing) is adjusted until the galvanometer reads zero. When the galvanometer reads zero, $\mathrm{emf}_{\mathrm{x}} = IR_{\mathrm{x}}$, where R_{x} is the resistance of the section of wire up to the contact point. Since no current flows through the

galvanometer, none flows through the unknown emf, and so emf_x is directly sensed.

Now, a very precisely known standard emf_s is substituted for emf_x, and the contact point is adjusted until the galvanometer again reads zero, so that $\text{emf}_\text{s} = IR_\text{s}$. In both cases, no current passes through the galvanometer, and so the current I through the long wire is the same. Upon taking the ratio $\frac{\text{emf}_\text{x}}{\text{emf}_\text{s}}$, I cancels, giving

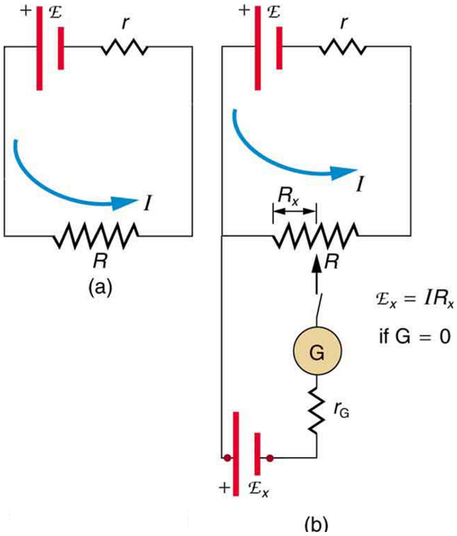
Equation:

$$\frac{\text{emf}_\text{x}}{\text{emf}_\text{s}} = \frac{IR_\text{x}}{IR_\text{s}} = \frac{R_\text{x}}{R_\text{s}}.$$

Solving for emf_x gives

Equation:

$$\text{emf}_\text{x} = \text{emf}_\text{s}\frac{R_\text{x}}{R_\text{s}}.$$



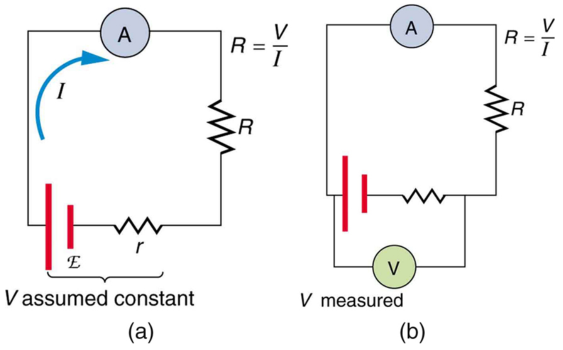
The potentiometer, a null measurement

device. (a) A voltage source connected to a long wire resistor passes a constant current I through it. (b) An unknown emf (labeled script E_{x} in the figure) is connected as shown, and the point of contact along R is adjusted until the galvanometer reads zero. The segment of wire has a resistance R_{x} and script $E_{\mathrm{x}} = IR_{\mathrm{x}}$, where I is unaffected by the connection since no current flows through the galvanometer. The unknown emf is thus proportional to the resistance of the wire segment.

Because a long uniform wire is used for R, the ratio of resistances $R_{\mathrm{x}}/R_{\mathrm{s}}$ is the same as the ratio of the lengths of wire that zero the galvanometer for each emf. The three quantities on the right-hand side of the equation are now known or measured, and $\mathrm{emf}_{\mathrm{x}}$ can be calculated. The uncertainty in this calculation can be considerably smaller than when using a voltmeter directly, but it is not zero. There is always some uncertainty in the ratio of resistances $R_{\mathrm{x}}/R_{\mathrm{s}}$ and in the standard $\mathrm{emf}_{\mathrm{s}}$. Furthermore, it is not possible to tell when the galvanometer reads exactly zero, which introduces error into both R_{x} and R_{s}, and may also affect the current I.

Resistance Measurements and the Wheatstone Bridge

There is a variety of so-called **ohmmeters** that purport to measure resistance. What the most common ohmmeters actually do is to apply a voltage to a resistance, measure the current, and calculate the resistance using Ohm's law. Their readout is this calculated resistance. Two configurations for ohmmeters using standard voltmeters and ammeters are shown in [link]. Such configurations are limited in accuracy, because the meters alter both the voltage applied to the resistor and the current that flows through it.



Two methods for measuring resistance with standard meters. (a) Assuming a known voltage for the source, an ammeter measures current, and resistance is calculated as $R = \frac{V}{I}$. (b) Since the terminal voltage V varies with current, it is better to measure it. V is most accurately known when I is small, but I itself is most accurately known when it is large.

The **Wheatstone bridge** is a null measurement device for calculating resistance by balancing potential drops in a circuit. (See [link].) The device is called a bridge because the galvanometer forms a bridge between two branches. A variety of **bridge devices** are used to make null measurements in circuits.

Resistors R_1 and R_2 are precisely known, while the arrow through R_3 indicates that it is a variable resistance. The value of R_3 can be precisely read. With the unknown resistance R_{x} in the circuit, R_3 is adjusted until the galvanometer reads zero. The potential difference between points b and d is then zero, meaning that b and d are at the same potential. With no current running through the galvanometer, it has no effect on the rest of the circuit. So the branches abc and adc are in parallel, and each branch has the full voltage of the source. That is, the IR drops along abc and adc are the same. Since b and d are at the same potential, the IR drop along ad must equal the IR drop along ab. Thus,

Equation:

$$I_1 R_1 = I_2 R_3.$$

Again, since b and d are at the same potential, the IR drop along dc must equal the IR drop along bc. Thus,

Equation:

$$I_1 R_2 = I_2 R_{\text{x}}.$$

Taking the ratio of these last two expressions gives

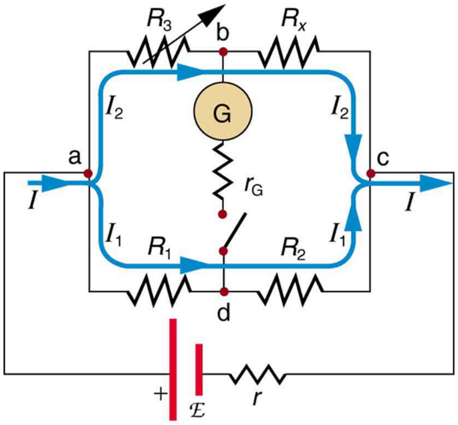
Equation:

$$\frac{I_1 R_1}{I_1 R_2} = \frac{I_2 R_3}{I_2 R_{\text{x}}}.$$

Canceling the currents and solving for R_x yields

Equation:

$$R_{\mathrm{x}} = R_3 \frac{R_2}{R_1}.$$



The Wheatstone bridge is used to calculate unknown resistances. The variable resistance R_3 is adjusted until the galvanometer reads zero with the switch closed. This simplifies the circuit, allowing R_{x} to be calculated based on the IR drops as discussed in the text.

This equation is used to calculate the unknown resistance when current through the galvanometer is zero. This method can be very accurate (often to four significant digits), but it is limited by two factors. First, it is not possible to get the current through the galvanometer to be exactly zero.

Second, there are always uncertainties in R_1, R_2, and R_3, which contribute to the uncertainty in R_{x}.

Exercise:

Check Your Understanding

Problem:

Identify other factors that might limit the accuracy of null measurements. Would the use of a digital device that is more sensitive than a galvanometer improve the accuracy of null measurements?

Solution:

One factor would be resistance in the wires and connections in a null measurement. These are impossible to make zero, and they can change over time. Another factor would be temperature variations in resistance, which can be reduced but not completely eliminated by choice of material. Digital devices sensitive to smaller currents than analog devices do improve the accuracy of null measurements because they allow you to get the current closer to zero.

Section Summary

- Null measurement techniques achieve greater accuracy by balancing a circuit so that no current flows through the measuring device.
- One such device, for determining voltage, is a potentiometer.
- Another null measurement device, for determining resistance, is the Wheatstone bridge.
- Other physical quantities can also be measured with null measurement techniques.

Conceptual questions

Exercise:

Problem:

Why can a null measurement be more accurate than one using standard voltmeters and ammeters? What factors limit the accuracy of null measurements?

Exercise:

Problem:

If a potentiometer is used to measure cell emfs on the order of a few volts, why is it most accurate for the standard emf_s to be the same order of magnitude and the resistances to be in the range of a few ohms?

Problem Exercises

Exercise:

Problem:

What is the emf_x of a cell being measured in a potentiometer, if the standard cell's emf is 12.0 V and the potentiometer balances for $R_\text{x} = 5.000\,\Omega$ and $R_\text{s} = 2.500\,\Omega$?

Solution:

24.0 V

Exercise:

Problem:

Calculate the emf_x of a dry cell for which a potentiometer is balanced when $R_\text{x} = 1.200\ \Omega$, while an alkaline standard cell with an emf of 1.600 V requires $R_\text{s} = 1.247\ \Omega$ to balance the potentiometer.

Exercise:

Problem:

When an unknown resistance R_x is placed in a Wheatstone bridge, it is possible to balance the bridge by adjusting R_3 to be $2500\ \Omega$. What is R_x if $\frac{R_2}{R_1} = 0.625$?

Solution:

$1.56\ \mathrm{k}\Omega$

Exercise:

Problem:

To what value must you adjust R_3 to balance a Wheatstone bridge, if the unknown resistance R_x is $100\,\Omega$, R_1 is $50.0\,\Omega$, and R_2 is $175\,\Omega$?

Exercise:

Problem:

(a) What is the unknown emf_x in a potentiometer that balances when R_x is $10.0\,\Omega$, and balances when R_s is $15.0\,\Omega$ for a standard 3.000-V emf? (b) The same emf_x is placed in the same potentiometer, which now balances when R_s is $15.0\,\Omega$ for a standard emf of 3.100 V. At what resistance R_x will the potentiometer balance?

Solution:

(a) 2.00 V

(b) 9.68 Ω

Exercise:

Problem:

Suppose you want to measure resistances in the range from $10.0\ \Omega$ to $10.0\ \mathrm{k}\Omega$ using a Wheatstone bridge that has $\frac{R_2}{R_1} = 2.000$. Over what range should R_3 be adjustable?

Solution:
Equation:

$$\text{Range} = 5.00\ \Omega \text{ to } 5.00\ \text{k}\Omega$$

Glossary

null measurements

methods of measuring current and voltage more accurately by balancing the circuit so that no current flows through the measurement device

potentiometer

a null measurement device for measuring potentials (voltages)

ohmmeter

an instrument that applies a voltage to a resistance, measures the current, calculates the resistance using Ohm’s law, and provides a readout of this calculated resistance

bridge device

a device that forms a bridge between two branches of a circuit; some bridge devices are used to make null measurements in circuits

Wheatstone bridge

a null measurement device for calculating resistance by balancing potential drops in a circuit

DC Circuits Containing Resistors and Capacitors

- Explain the importance of the time constant, τ , and calculate the time constant for a given resistance and capacitance.
- Explain why batteries in a flashlight gradually lose power and the light dims over time.
- Describe what happens to a graph of the voltage across a capacitor over time as it charges.
- Explain how a timing circuit works and list some applications.
- Calculate the necessary speed of a strobe flash needed to “stop” the movement of an object over a particular length.

When you use a flash camera, it takes a few seconds to charge the capacitor that powers the flash. The light flash discharges the capacitor in a tiny fraction of a second. Why does charging take longer than discharging? This question and a number of other phenomena that involve charging and discharging capacitors are discussed in this module.

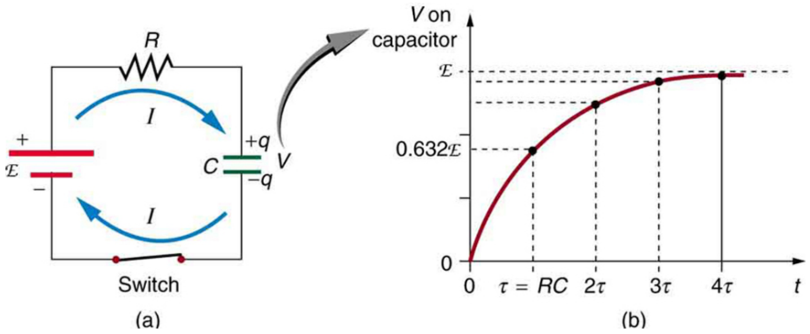
RC Circuits

An **$\mathbf{RC}$ circuit** is one containing a **resistor** R and a **capacitor** C. The capacitor is an electrical component that stores electric charge.

[link] shows a simple RC circuit that employs a DC (direct current) voltage source. The capacitor is initially uncharged. As soon as the switch is closed, current flows to and from the initially uncharged capacitor. As charge increases on the capacitor plates, there is increasing opposition to the flow of charge by the repulsion of like charges on each plate.

In terms of voltage, this is because voltage across the capacitor is given by $V_{\text{c}} = Q/C$, where Q is the amount of charge stored on each plate and C is the **capacitance**. This voltage opposes the battery, growing from zero to the maximum emf when fully charged. The current thus decreases from its initial value of $I_0 = \frac{\text{emf}}{R}$ to zero as the voltage on the capacitor reaches the same value as the emf. When there is no current, there is no IR drop, and so the voltage on the capacitor must then equal the emf of the voltage source. This can also be explained with Kirchhoff’s second rule (the loop rule), discussed in Kirchhoff’s Rules, which says that the algebraic sum of changes in potential around any closed loop must be zero.

The initial current is $I_0 = \frac{\text{emf}}{R}$, because all of the IR drop is in the resistance. Therefore, the smaller the resistance, the faster a given capacitor will be charged. Note that the internal resistance of the voltage source is included in R, as are the resistances of the capacitor and the connecting wires. In the flash camera scenario above, when the batteries powering the camera begin to wear out, their internal resistance rises, reducing the current and lengthening the time it takes to get ready for the next flash.



(a) An RC circuit with an initially uncharged capacitor. Current flows in the direction shown (opposite of electron flow) as soon as the switch is closed. Mutual repulsion of like charges in the capacitor progressively slows the flow as the capacitor is charged, stopping the current when the capacitor is fully charged and $Q = C \cdot \text{emf}$. (b) A graph of voltage across the capacitor versus time, with the switch closing at time $t = 0$. (Note that in the two parts of the figure, the capital script E stands for emf, q stands for the charge stored on the capacitor, and τ is the RC time constant.)

Voltage on the capacitor is initially zero and rises rapidly at first, since the initial current is a maximum. [link](b) shows a graph of capacitor voltage versus time (t) starting when the switch is closed at $t=0$. The voltage approaches emf asymptotically, since the closer it gets to emf the less current flows. The equation for voltage versus time when charging a capacitor C through a resistor R, derived using calculus, is

Equation:

$$V=\text{emf}(1-e^{-t/\text{RC}})\text{ (charging)},$$

where V is the voltage across the capacitor, emf is equal to the emf of the DC voltage source, and the exponential e = 2.718 … is the base of the natural logarithm. Note that the units of RC are seconds. We define

Equation:

$$\tau=\text{RC},$$

where τ (the Greek letter tau) is called the time constant for an RC circuit. As noted before, a small resistance R allows the capacitor to charge faster. This is reasonable, since a larger current flows through a smaller resistance. It is also reasonable that the smaller the capacitor C, the less time needed to charge it. Both factors are contained in $\tau=\text{RC}$.

More quantitatively, consider what happens when $t=\tau=\text{RC}$. Then the voltage on the capacitor is

Equation:

$$V=\text{emf}\left(1-e^{-1}\right)=\text{emf}(1-0.368)=0.632\cdot\text{emf}.$$

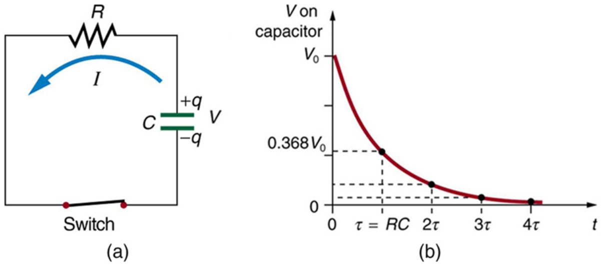
This means that in the time $\tau=\text{RC}$, the voltage rises to 0.632 of its final value. The voltage will rise 0.632 of the remainder in the next time τ. It is a characteristic of the exponential function that the final value is never reached, but 0.632 of the remainder to that value is achieved in every time, τ. In just a few multiples of the time constant τ, then, the final value is very nearly achieved, as the graph in [link](b) illustrates.

Discharging a Capacitor

Discharging a capacitor through a resistor proceeds in a similar fashion, as [link] illustrates. Initially, the current is $I_0 = \frac{V_0}{R}$, driven by the initial voltage V_0 on the capacitor. As the voltage decreases, the current and hence the rate of discharge decreases, implying another exponential formula for V . Using calculus, the voltage V on a capacitor C being discharged through a resistor R is found to be

Equation:

$$V = V_0 \, e^{-t/\text{RC}}(\text{discharging}).$$



(a) Closing the switch discharges the capacitor C through the resistor R. Mutual repulsion of like charges on each plate drives the current. (b) A graph of voltage across the capacitor versus time, with $V = V_0$ at $t = 0$. The voltage decreases exponentially, falling a fixed fraction of the way to zero in each subsequent time constant τ.

The graph in [link](b) is an example of this exponential decay. Again, the time constant is $\tau = \text{RC}$. A small resistance R allows the capacitor to discharge in a small time, since the current is larger. Similarly, a small capacitance requires less time to discharge, since less charge is stored. In the first time interval $\tau = \text{RC}$ after the switch is closed, the voltage falls to 0.368 of its initial value, since $V = V_0 \cdot e^{-1} = 0.368V_0$.

During each successive time τ, the voltage falls to 0.368 of its preceding value. In a few multiples of τ, the voltage becomes very close to zero, as indicated by the graph in [link](b).

Now we can explain why the flash camera in our scenario takes so much longer to charge than discharge; the resistance while charging is significantly greater than while discharging. The internal resistance of the battery accounts for most of the resistance while charging. As the battery ages, the increasing internal resistance makes the charging process even slower. (You may have noticed this.)

The flash discharge is through a low-resistance ionized gas in the flash tube and proceeds very rapidly. Flash photographs, such as in [link], can capture a brief instant of a rapid motion because the flash can be less than a microsecond in duration. Such flashes can be made extremely intense.

During World War II, nighttime reconnaissance photographs were made from the air with a single flash illuminating more than a square kilometer of enemy territory. The brevity of the flash eliminated blurring due to the surveillance aircraft’s motion. Today, an important use of intense flash lamps is to pump energy into a laser. The short intense flash can rapidly energize a laser and allow it to reemit the energy in another form.

This stop-motion photograph of a rufous hummingbird (*Selasphorus rufus*) feeding on a flower was obtained with an extremely brief and intense flash of light powered by the discharge of a capacitor through a gas. (credit: Dean E. Biggins, U.S. Fish and Wildlife Service)

Example:

Integrated Concept Problem: Calculating Capacitor Size—Strobe Lights

High-speed flash photography was pioneered by Doc Edgerton in the 1930s, while he was a professor of electrical engineering at MIT. You might have seen examples of his work in the amazing shots of hummingbirds in motion, a drop of milk splattering on a table, or a bullet penetrating an apple (see [link]). To stop the motion and capture these pictures, one needs a high-intensity, very short pulsed flash, as mentioned earlier in this module.

Suppose one wished to capture the picture of a bullet (moving at $5.0 \times 10^2\ \mathrm{m/s}$) that was passing through an apple. The duration of the flash is related to the RC time constant, τ. What size capacitor would one need in the RC circuit to succeed, if the resistance of the flash tube was $10.0\ \Omega$? Assume the apple is a sphere with a diameter of 8.0×10^{-2} m.

Strategy

We begin by identifying the physical principles involved. This example deals with the strobe light, as discussed above. [link] shows the circuit for this probe. The characteristic time τ of the strobe is given as $\tau = \mathrm{RC}$.

Solution

We wish to find C, but we don't know τ. We want the flash to be on only while the bullet traverses the apple. So we need to use the kinematic equations that describe the relationship between distance x, velocity v, and time t:

Equation:

$$x = \mathrm{vt} \text{ or } t = \frac{x}{v}.$$

The bullet's velocity is given as $5.0 \times 10^{2}\ \mathrm{m/s}$, and the distance x is 8.0×10^{-2} m. The traverse time, then, is

Equation:

$$t = \frac{x}{v} = \frac{8.0 \times 10^{-2}\ \mathrm{m}}{5.0 \times 10^{2}\ \mathrm{m/s}} = 1.6 \times 10^{-4}\ \mathrm{s}.$$

We set this value for the crossing time t equal to τ. Therefore,

Equation:

$$C = \frac{t}{R} = \frac{1.6{\times}10^{-4}\ \mathrm{s}}{10.0\ \Omega} = 16\ \mu\mathrm{F}.$$

(Note: Capacitance C is typically measured in farads, F, defined as Coulombs per volt. From the equation, we see that C can also be stated in units of seconds per ohm.)

Discussion

The flash interval of 160 μs (the traverse time of the bullet) is relatively easy to obtain today. Strobe lights have opened up new worlds from science to entertainment. The information from the picture of the apple and bullet was used in the Warren Commission Report on the assassination of

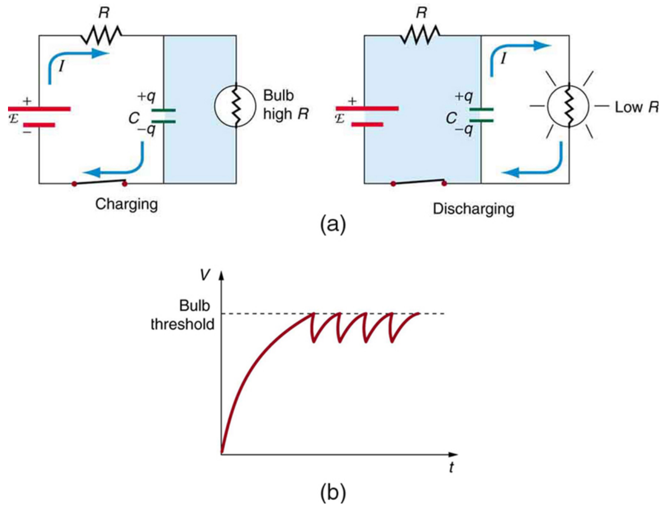
President John F. Kennedy in 1963 to confirm that only one bullet was fired.

RC Circuits for Timing

RC circuits are commonly used for timing purposes. A mundane example of this is found in the ubiquitous intermittent wiper systems of modern cars. The time between wipes is varied by adjusting the resistance in an RC circuit. Another example of an RC circuit is found in novelty jewelry, Halloween costumes, and various toys that have battery-powered flashing lights. (See [link] for a timing circuit.)

A more crucial use of RC circuits for timing purposes is in the artificial pacemaker, used to control heart rate. The heart rate is normally controlled by electrical signals generated by the sino-atrial (SA) node, which is on the wall of the right atrium chamber. This causes the muscles to contract and pump blood. Sometimes the heart rhythm is abnormal and the heartbeat is too high or too low.

The artificial pacemaker is inserted near the heart to provide electrical signals to the heart when needed with the appropriate time constant. Pacemakers have sensors that detect body motion and breathing to increase the heart rate during exercise to meet the body’s increased needs for blood and oxygen.



(a) The lamp in this RC circuit ordinarily has a very high resistance, so that the battery charges the capacitor as if the lamp were not there. When the voltage reaches a threshold value, a current flows through the lamp that dramatically reduces its resistance, and the capacitor discharges through the lamp as if the battery and charging resistor were not there. Once discharged, the process starts again, with the flash period determined by the RC constant τ. (b) A graph of voltage versus time for this circuit.

Example:

Calculating Time: *RC* Circuit in a Heart Defibrillator

A heart defibrillator is used to resuscitate an accident victim by discharging a capacitor through the trunk of her body. A simplified version of the circuit is seen in [link]. (a) What is the time constant if an 8.00-μF capacitor is used and the path resistance through her body is $1.00 \times 10^3\ \Omega$? (b) If the initial voltage is 10.0 kV, how long does it take to decline to 5.00×10^2 V?

Strategy

Since the resistance and capacitance are given, it is straightforward to multiply them to give the time constant asked for in part (a). To find the time for the voltage to decline to 5.00×10^2 V, we repeatedly multiply the initial voltage by 0.368 until a voltage less than or equal to 5.00×10^2 V is obtained. Each multiplication corresponds to a time of τ seconds.

Solution for (a)

The time constant τ is given by the equation $\tau = \text{RC}$. Entering the given values for resistance and capacitance (and remembering that units for a farad can be expressed as s/Ω) gives

Equation:

$$\tau = \text{RC} = (1.00 \times 10^3\,\Omega)(8.00\ \mu F) = 8.00\ \text{ms.}$$

Solution for (b)

In the first 8.00 ms, the voltage (10.0 kV) declines to 0.368 of its initial value. That is:

Equation:

$$V = 0.368 V_0 = 3.680 \times 10^3\ \text{V at } t = 8.00\ \text{ms.}$$

(Notice that we carry an extra digit for each intermediate calculation.) After another 8.00 ms, we multiply by 0.368 again, and the voltage is

Equation:

$$\begin{array}{rcl} V\prime & = & 0.368 V \\ & = & (0.368)\left(3.680 \times 10^3\ \text{V}\right) \\ & = & 1.354 \times 10^3\ \text{V at } t = 16.0\ \text{ms.} \end{array}$$

Similarly, after another 8.00 ms, the voltage is

Equation:

$$\begin{aligned} V'' &= 0.368V' = (0.368)(1.354 \times 10^3 \text{ V}) \\ &= 498 \text{ V at } t = 24.0 \text{ ms.} \end{aligned}$$

Discussion

So after only 24.0 ms, the voltage is down to 498 V, or 4.98% of its original value.Such brief times are useful in heart defibrillation, because the brief but intense current causes a brief but effective contraction of the heart. The actual circuit in a heart defibrillator is slightly more complex than the one in [link], to compensate for magnetic and AC effects that will be covered in Magnetism.

Exercise:
Check Your Understanding

Problem: When is the potential difference across a capacitor an emf?

Solution:

Only when the current being drawn from or put into the capacitor is zero. Capacitors, like batteries, have internal resistance, so their output voltage is not an emf unless current is zero. This is difficult to measure in practice so we refer to a capacitor’s voltage rather than its emf. But the source of potential difference in a capacitor is fundamental and it is an emf.

Note:

PhET Explorations: Circuit Construction Kit (DC only)

An electronics kit in your computer! Build circuits with resistors, light bulbs, batteries, and switches. Take measurements with the realistic ammeter and voltmeter. View the circuit as a schematic diagram, or switch to a life-like view.

https://archive.cnx.org/specials/f23ce496-c9d1-11e5-bdc8-bb04dc1eecb6/circuit-construction-kit-dc-only/#sim-cck

Section Summary

- An RC circuit is one that has both a resistor and a capacitor.
- The time constant τ for an RC circuit is $\tau = \mathrm{RC}$.
- When an initially uncharged ($V_0 = 0$ at $t = 0$) capacitor in series with a resistor is charged by a DC voltage source, the voltage rises, asymptotically approaching the emf of the voltage source; as a function of time,
  Equation:
  $$V = \mathrm{emf}(1 - e^{-t/\mathrm{RC}})(\mathrm{charging}).$$
- Within the span of each time constant τ, the voltage rises by 0.632 of the remaining value, approaching the final voltage asymptotically.
- If a capacitor with an initial voltage V_0 is discharged through a resistor starting at $t = 0$, then its voltage decreases exponentially as given by
  Equation:
  $$V = V_0 e^{-t/\mathrm{RC}}(\mathrm{discharging}).$$
- In each time constant τ, the voltage falls by 0.368 of its remaining initial value, approaching zero asymptotically.

Conceptual questions

Exercise:

Problem:

Regarding the units involved in the relationship $\tau = \mathrm{RC}$, verify that the units of resistance times capacitance are time, that is, $\Omega \cdot \mathrm{F} = \mathrm{s}$.

Exercise:

Problem:

The RC time constant in heart defibrillation is crucial to limiting the time the current flows. If the capacitance in the defibrillation unit is fixed, how would you manipulate resistance in the circuit to adjust the RC constant τ? Would an adjustment of the applied voltage also be needed to ensure that the current delivered has an appropriate value?

Exercise:

Problem:

When making an ECG measurement, it is important to measure voltage variations over small time intervals. The time is limited by the RC constant of the circuit—it is not possible to measure time variations shorter than RC. How would you manipulate R and C in the circuit to allow the necessary measurements?

Exercise:

Problem:

Draw two graphs of charge versus time on a capacitor. Draw one for charging an initially uncharged capacitor in series with a resistor, as in the circuit in [link], starting from $\mathrm{t} = 0$. Draw the other for discharging a capacitor through a resistor, as in the circuit in [link], starting at $\mathrm{t} = 0$, with an initial charge Q_0. Show at least two intervals of τ.

Exercise:

Problem:

When charging a capacitor, as discussed in conjunction with [link], how long does it take for the voltage on the capacitor to reach emf? Is this a problem?

Exercise:

Problem:

When discharging a capacitor, as discussed in conjunction with [link], how long does it take for the voltage on the capacitor to reach zero? Is this a problem?

Exercise:

Problem:

Referring to [link], draw a graph of potential difference across the resistor versus time, showing at least two intervals of τ. Also draw a graph of current versus time for this situation.

Exercise:

Problem:

A long, inexpensive extension cord is connected from inside the house to a refrigerator outside. The refrigerator doesn't run as it should. What might be the problem?

Exercise:

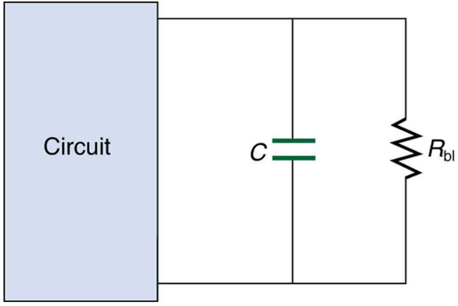
Problem:

In [link], does the graph indicate the time constant is shorter for discharging than for charging? Would you expect ionized gas to have low resistance? How would you adjust R to get a longer time between flashes? Would adjusting R affect the discharge time?

Exercise:

Problem:

An electronic apparatus may have large capacitors at high voltage in the power supply section, presenting a shock hazard even when the apparatus is switched off. A “bleeder resistor” is therefore placed across such a capacitor, as shown schematically in [link], to bleed the charge from it after the apparatus is off. Why must the bleeder resistance be much greater than the effective resistance of the rest of the circuit? How does this affect the time constant for discharging the capacitor?



A bleeder resistor R_{bl} discharges the capacitor in this electronic device once it is switched off.

Problem Exercises

Exercise:

Problem:

The timing device in an automobile’s intermittent wiper system is based on an RC time constant and utilizes a $0.500\text{-}\mu\text{F}$ capacitor and a variable resistor. Over what range must R be made to vary to achieve time constants from 2.00 to 15.0 s?

Solution:

range $4.00 \text{ to } 30.0 \text{ M}\Omega$

Exercise:

Problem:

A heart pacemaker fires 72 times a minute, each time a 25.0-nF capacitor is charged (by a battery in series with a resistor) to 0.632 of its full voltage. What is the value of the resistance?

Exercise:

Problem:

The duration of a photographic flash is related to an RC time constant, which is $0.100\ \mu\text{s}$ for a certain camera. (a) If the resistance of the flash lamp is $0.0400\,\Omega$ during discharge, what is the size of the capacitor supplying its energy? (b) What is the time constant for charging the capacitor, if the charging resistance is $800\ \text{k}\Omega$?

Solution:

(a) $2.50\ \mu\text{F}$

(b) 2.00 s

Exercise:

Problem:

A 2.00- and a $7.50\text{-}\mu\text{F}$ capacitor can be connected in series or parallel, as can a 25.0- and a $100\text{-k}\Omega$ resistor. Calculate the four RC time constants possible from connecting the resulting capacitance and resistance in series.

Exercise:

Problem:

After two time constants, what percentage of the final voltage, emf, is on an initially uncharged capacitor C, charged through a resistance R?

Solution:

86.5%

Exercise:

Problem:

A $500\text{-}\Omega$ resistor, an uncharged $1.50\text{-}\mu\text{F}$ capacitor, and a 6.16-V emf are connected in series. (a) What is the initial current? (b) What is the RC time constant? (c) What is the current after one time constant? (d) What is the voltage on the capacitor after one time constant?

Exercise:

Problem:

A heart defibrillator being used on a patient has an RC time constant of 10.0 ms due to the resistance of the patient and the capacitance of the defibrillator. (a) If the defibrillator has an $8.00\text{-}\mu\text{F}$ capacitance, what is the resistance of the path through the patient? (You may neglect the capacitance of the patient and the resistance of the defibrillator.) (b) If the initial voltage is 12.0 kV, how long does it take to decline to 6.00×10^2 V?

Solution:

(a) $1.25\ \text{k}\Omega$

(b) 30.0 ms

Exercise:

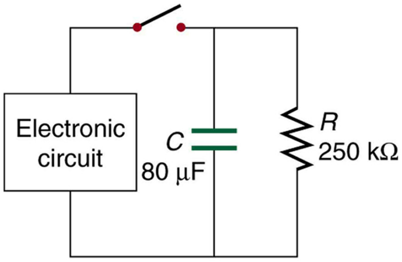
Problem:

An ECG monitor must have an RC time constant less than 1.00×10^2 μs to be able to measure variations in voltage over small time intervals. (a) If the resistance of the circuit (due mostly to that of the patient's chest) is $1.00\ \text{k}\Omega$, what is the maximum capacitance of the circuit? (b) Would it be difficult in practice to limit the capacitance to less than the value found in (a)?

Exercise:

Problem:

[link] shows how a bleeder resistor is used to discharge a capacitor after an electronic device is shut off, allowing a person to work on the electronics with less risk of shock. (a) What is the time constant? (b) How long will it take to reduce the voltage on the capacitor to 0.250% (5% of 5%) of its full value once discharge begins? (c) If the capacitor is charged to a voltage V_0 through a $100\text{-}\Omega$ resistance, calculate the time it takes to rise to $0.865V_0$ (This is about two time constants.)



Solution:

(a) 20.0 s

(b) 120 s

(c) 16.0 ms

Exercise:

Problem:

Using the exact exponential treatment, find how much time is required to discharge a $250\text{-}\mu\text{F}$ capacitor through a $500\text{-}\Omega$ resistor down to 1.00% of its original voltage.

Exercise:

Problem:

Using the exact exponential treatment, find how much time is required to charge an initially uncharged 100-pF capacitor through a $75.0\text{-M}\Omega$ resistor to 90.0% of its final voltage.

Solution:

1.73×10^{-2} s

Exercise:

Problem: Integrated Concepts

If you wish to take a picture of a bullet traveling at 500 m/s, then a very brief flash of light produced by an RC discharge through a flash tube can limit blurring. Assuming 1.00 mm of motion during one RC constant is acceptable, and given that the flash is driven by a $600\text{-}\mu\text{F}$ capacitor, what is the resistance in the flash tube?

Solution:

$3.33 \times 10^{-3}\ \Omega$

Exercise:

Problem: Integrated Concepts

A flashing lamp in a Christmas earring is based on an RC discharge of a capacitor through its resistance. The effective duration of the flash is 0.250 s, during which it produces an average 0.500 W from an average

3.00 V. (a) What energy does it dissipate? (b) How much charge moves through the lamp? (c) Find the capacitance. (d) What is the resistance of the lamp?

Exercise:

Problem: Integrated Concepts

A 160-μF capacitor charged to 450 V is discharged through a $31.2\text{-k}\Omega$ resistor. (a) Find the time constant. (b) Calculate the temperature increase of the resistor, given that its mass is 2.50 g and its specific heat is $1.67\frac{\text{kJ}}{\text{kg}\cdot{}^{\text{o}}\text{C}}$, noting that most of the thermal energy is retained in the short time of the discharge. (c) Calculate the new resistance, assuming it is pure carbon. (d) Does this change in resistance seem significant?

Solution:

(a) 4.99 s

(b) 3.87^{o}C

(c) $31.1\ \text{k}\Omega$

(d) No

Exercise:

Problem: Unreasonable Results

(a) Calculate the capacitance needed to get an RC time constant of 1.00×10^{3} s with a $0.100\text{-}\Omega$ resistor. (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem: Construct Your Own Problem

Consider a camera’s flash unit. Construct a problem in which you calculate the size of the capacitor that stores energy for the flash lamp. Among the things to be considered are the voltage applied to the capacitor, the energy needed in the flash and the associated charge needed on the capacitor, the resistance of the flash lamp during discharge, and the desired RC time constant.

Exercise:

Problem: Construct Your Own Problem

Consider a rechargeable lithium cell that is to be used to power a camcorder. Construct a problem in which you calculate the internal resistance of the cell during normal operation. Also, calculate the minimum voltage output of a battery charger to be used to recharge your lithium cell. Among the things to be considered are the emf and useful terminal voltage of a lithium cell and the current it should be able to supply to a camcorder.

Glossary

RC circuit	a circuit that contains both a resistor and a capacitor
capacitor	an electrical component used to store energy by separating electric charge on two opposing plates
capacitance	the maximum amount of electric potential energy that can be stored (or separated) for a given electric potential

Introduction to Magnetism class="introduction"

The magnificent spectacle of the Aurora Borealis, or northern lights, glows in the northern sky above Bear Lake near Eielson Air Force Base, Alaska. Shaped by the Earth’s magnetic field, this light is produced by radiation spewed from solar storms. (credit: Senior Airman Joshua Strang, via Flickr)

One evening, an Alaskan sticks a note to his refrigerator with a small magnet. Through the kitchen window, the Aurora Borealis glows in the night sky. This grand spectacle is shaped by the same force that holds the note to the refrigerator.

People have been aware of magnets and magnetism for thousands of years. The earliest records date to well before the time of Christ, particularly in a region of Asia Minor called Magnesia (the name of this region is the source of words like *magnetic*). Magnetic rocks found in Magnesia, which is now part of western Turkey, stimulated interest during ancient times. A practical application for magnets was found later, when they were employed as navigational compasses. The use of magnets in compasses resulted not only in improved long-distance sailing, but also in the names of “north” and “south” being given to the two types of magnetic poles.

Today magnetism plays many important roles in our lives. Physicists’ understanding of magnetism has enabled the development of technologies that affect our everyday lives. The iPod in your purse or backpack, for example, wouldn’t have been possible without the applications of magnetism and electricity on a small scale.

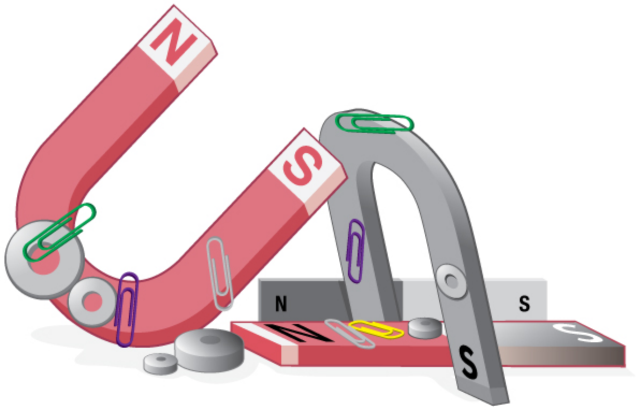
The discovery that weak changes in a magnetic field in a thin film of iron and chromium could bring about much larger changes in electrical resistance was one of the first large successes of nanotechnology. The 2007 Nobel Prize in Physics went to Albert Fert from France and Peter Grunberg from Germany for this discovery of *giant magnetoresistance* and its applications to computer memory.

All electric motors, with uses as diverse as powering refrigerators, starting cars, and moving elevators, contain magnets. Generators, whether producing hydroelectric power or running bicycle lights, use magnetic fields. Recycling facilities employ magnets to separate iron from other refuse. Hundreds of millions of dollars are spent annually on magnetic containment of fusion as a future energy source. Magnetic resonance imaging (MRI) has become an important diagnostic tool in the field of medicine, and the use of magnetism to explore brain activity is a subject of contemporary research and development. The list of applications also includes computer hard drives, tape recording, detection of inhaled asbestos, and levitation of high-speed trains. Magnetism is used to explain atomic energy levels, cosmic rays, and charged particles trapped in the Van Allen belts. Once again, we will find all these disparate phenomena are linked by a small number of underlying physical principles.

Engineering of technology like iPods would not be possible without a deep understanding magnetism. (credit: Jesse! S?, Flickr)

Magnets

- Describe the difference between the north and south poles of a magnet.
- Describe how magnetic poles interact with each other.

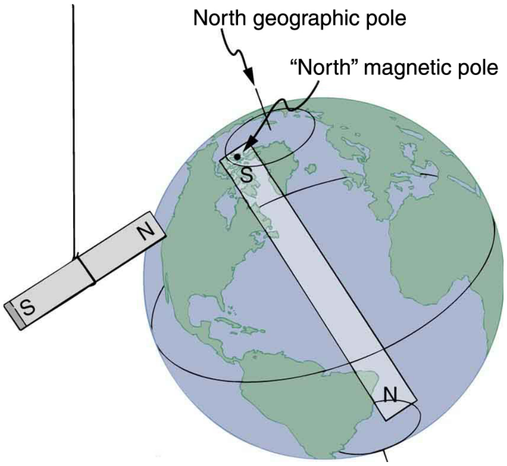


Magnets come in various shapes, sizes, and strengths. All have both a north pole and a south pole. There is never an isolated pole (a monopole).

All magnets attract iron, such as that in a refrigerator door. However, magnets may attract or repel other magnets. Experimentation shows that all magnets have two poles. If freely suspended, one pole will point toward the north. The two poles are thus named the **north magnetic pole** and the **south magnetic pole** (or more properly, north-seeking and south-seeking poles, for the attractions in those directions).

Note:
Universal Characteristics of Magnets and Magnetic Poles
It is a universal characteristic of all magnets that *like poles repel and unlike poles attract*. (Note the similarity with electrostatics: unlike charges attract and like charges repel.)

Further experimentation shows that it is *impossible to separate north and south poles* in the manner that + and − charges can be separated.



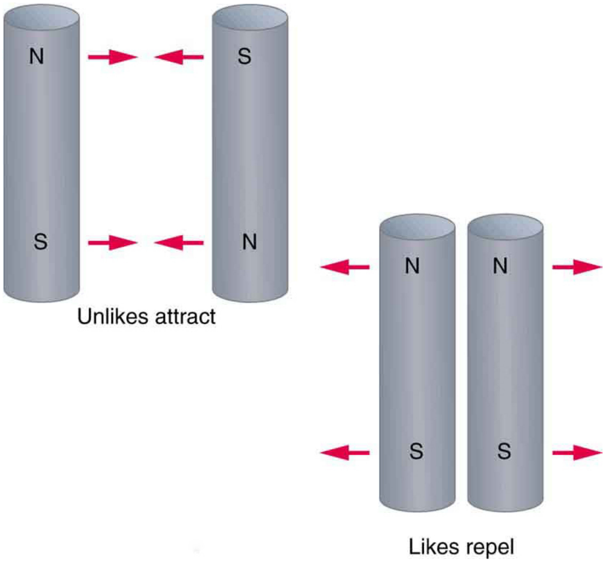
One end of a bar magnet is suspended from a thread that points toward north. The magnet's two poles are labeled N and S for north-seeking and south-seeking poles, respectively.

Note:

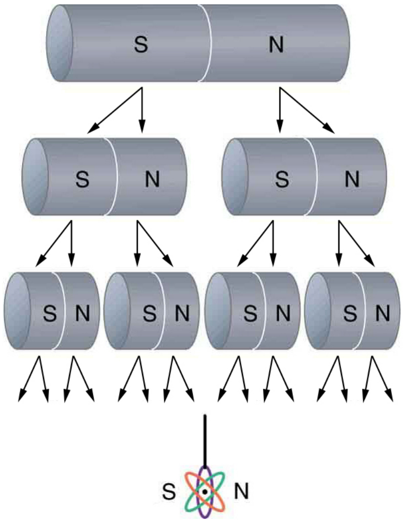
Misconception Alert: Earth's Geographic North Pole Hides an S

The Earth acts like a very large bar magnet with its south-seeking pole near the geographic North Pole. That is why the north pole of your compass is attracted toward the geographic north pole of the Earth—because the magnetic pole that is near the geographic North Pole is actually a south magnetic pole! Confusion arises because the geographic term "North Pole" has come to be used (incorrectly) for the magnetic pole that is near the

North Pole. Thus, “North magnetic pole” is actually a misnomer—it should be called the South magnetic pole.


N
S
S
N
Unlikes attract
N
N
S
S
Likes repel

Unlike poles attract, whereas like poles repel.


S
N
S
N
S
N
S N
S N
S N
S N
S
N

North and south poles always occur in pairs. Attempts to separate them result in more pairs of poles. If we continue to split the magnet, we will eventually get down to an iron atom with a north pole and a south pole—these, too, cannot be separated.

The fact that magnetic poles always occur in pairs of north and south is true from the very large scale—for example, sunspots always occur in pairs that are north and south magnetic poles—all the way down to the very small scale. Magnetic atoms have both a north pole and a south pole, as do many types of subatomic particles, such as electrons, protons, and neutrons.

Note:

Making Connections: Take-Home Experiment—Refrigerator Magnets

We know that like magnetic poles repel and unlike poles attract. See if you can show this for two refrigerator magnets. Will the magnets stick if you turn them over? Why do they stick to the door anyway? What can you say about the magnetic properties of the door next to the magnet? Do refrigerator magnets stick to metal or plastic spoons? Do they stick to all types of metal?

Section Summary

- Magnetism is a subject that includes the properties of magnets, the effect of the magnetic force on moving charges and currents, and the

creation of magnetic fields by currents.
- There are two types of magnetic poles, called the north magnetic pole and south magnetic pole.
- North magnetic poles are those that are attracted toward the Earth’s geographic north pole.
- Like poles repel and unlike poles attract.
- Magnetic poles always occur in pairs of north and south—it is not possible to isolate north and south poles.

Conceptual Questions

Exercise:

Problem:

Volcanic and other such activity at the mid-Atlantic ridge extrudes material to fill the gap between separating tectonic plates associated with continental drift. The magnetization of rocks is found to reverse in a coordinated manner with distance from the ridge. What does this imply about the Earth’s magnetic field and how could the knowledge of the spreading rate be used to give its historical record?

Glossary

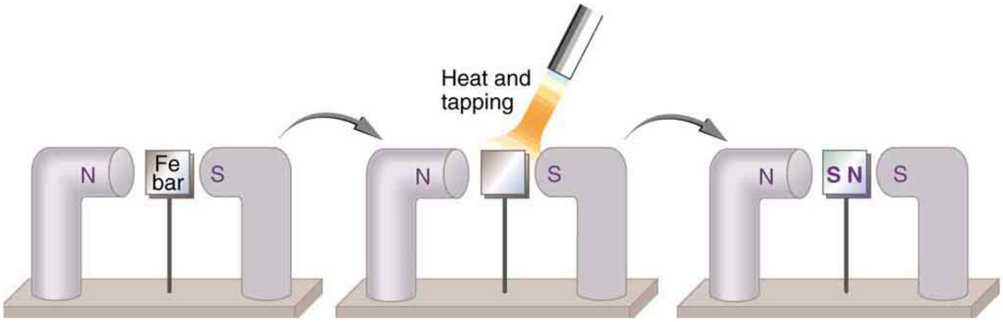
north magnetic pole	the end or the side of a magnet that is attracted toward Earth’s geographic north pole
south magnetic pole	the end or the side of a magnet that is attracted toward Earth’s geographic south pole

Ferromagnets and Electromagnets

- Define ferromagnet.
- Describe the role of magnetic domains in magnetization.
- Explain the significance of the Curie temperature.
- Describe the relationship between electricity and magnetism.

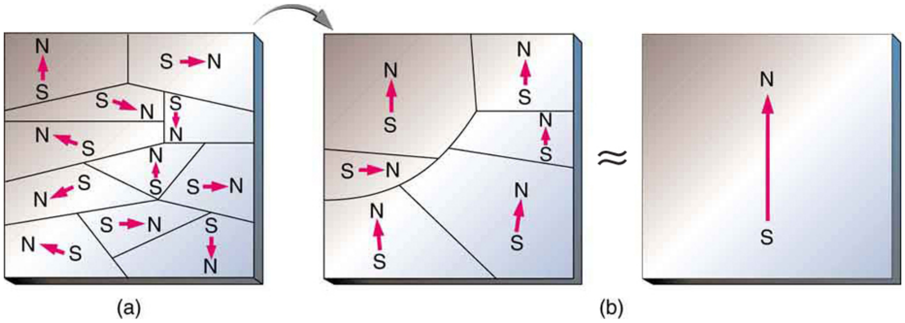
Ferromagnets

Only certain materials, such as iron, cobalt, nickel, and gadolinium, exhibit strong magnetic effects. Such materials are called **ferromagnetic**, after the Latin word for iron, *ferrum*. A group of materials made from the alloys of the rare earth elements are also used as strong and permanent magnets; a popular one is neodymium. Other materials exhibit weak magnetic effects, which are detectable only with sensitive instruments. Not only do ferromagnetic materials respond strongly to magnets (the way iron is attracted to magnets), they can also be **magnetized** themselves—that is, they can be induced to be magnetic or made into permanent magnets.



An unmagnetized piece of iron is placed between two magnets, heated, and then cooled, or simply tapped when cold. The iron becomes a permanent magnet with the poles aligned as shown: its south pole is adjacent to the north pole of the original magnet, and its north pole is adjacent to the south pole of the original magnet. Note that there are attractive forces between the magnets.

When a magnet is brought near a previously unmagnetized ferromagnetic material, it causes local magnetization of the material with unlike poles closest, as in [link]. (This results in the attraction of the previously unmagnetized material to the magnet.) What happens on a microscopic scale is illustrated in [link]. The regions within the material called **domains** act like small bar magnets. Within domains, the poles of individual atoms are aligned. Each atom acts like a tiny bar magnet. Domains are small and randomly oriented in an unmagnetized ferromagnetic object. In response to an external magnetic field, the domains may grow to millimeter size, aligning themselves as shown in [link](b). This induced magnetization can be made permanent if the material is heated and then cooled, or simply tapped in the presence of other magnets.



(a) An unmagnetized piece of iron (or other ferromagnetic material) has randomly oriented domains. (b) When magnetized by an external field, the domains show greater alignment, and some grow at the expense of others. Individual atoms are aligned within domains; each atom acts like a tiny bar magnet.

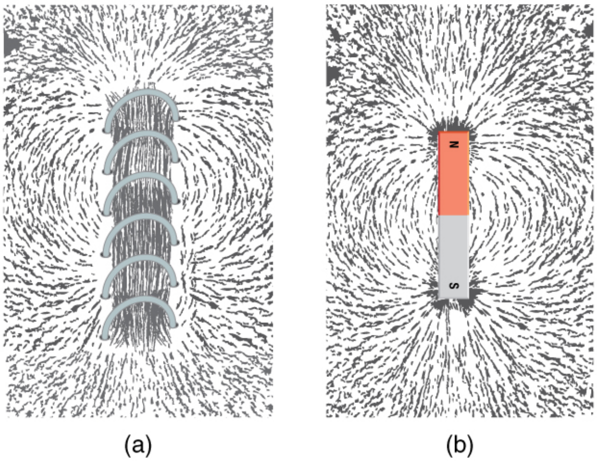
Conversely, a permanent magnet can be demagnetized by hard blows or by heating it in the absence of another magnet. Increased thermal motion at higher temperature can disrupt and randomize the orientation and the size of the domains. There is a well-defined temperature for ferromagnetic materials, which is called the **Curie temperature**, above which they cannot be magnetized. The Curie temperature for iron is 1043 K (770°C), which is well above room temperature. There are several elements and alloys that have Curie temperatures much lower than room temperature and are ferromagnetic only below those temperatures.

Electromagnets

Early in the 19th century, it was discovered that electrical currents cause magnetic effects. The first significant observation was by the Danish scientist Hans Christian Oersted (1777–1851), who found that a compass needle was deflected by a current-carrying wire. This was the first significant evidence that the movement of charges had any connection with magnets. **Electromagnetism** is the use of electric current to make magnets. These temporarily induced magnets are called **electromagnets**. Electromagnets are employed for everything from a wrecking yard crane that lifts scrapped cars to controlling the beam of a 90-km-circumference particle accelerator to the magnets in medical imaging machines (See [link]).

Instrument for magnetic resonance imaging (MRI). The device uses a superconducting cylindrical coil for the main magnetic field. The patient goes into this “tunnel” on the gurney. (credit: Bill McChesney, Flickr)

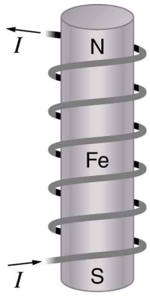
[link] shows that the response of iron filings to a current-carrying coil and to a permanent bar magnet. The patterns are similar. In fact, electromagnets and ferromagnets have the same basic characteristics—for example, they have north and south poles that cannot be separated and for which like poles repel and unlike poles attract.



(a) (b)

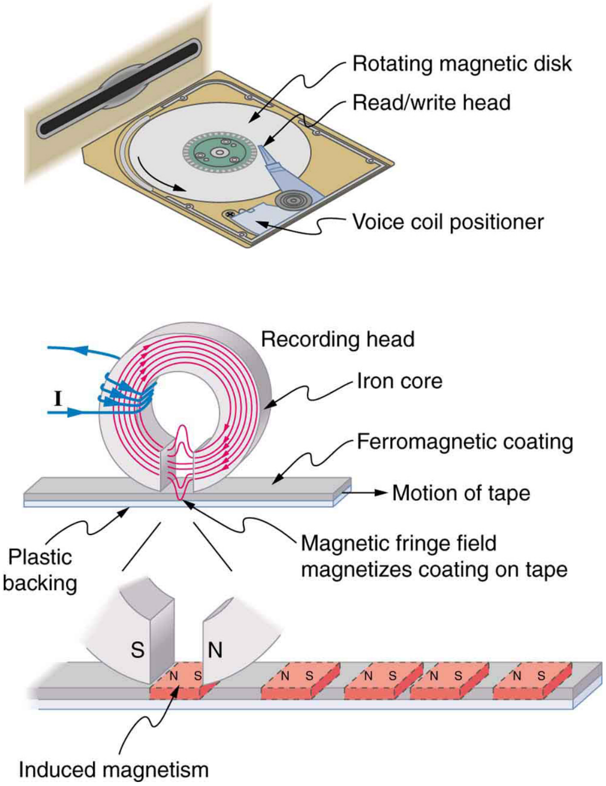
Iron filings near (a) a current-carrying coil and (b) a magnet act like tiny compass needles, showing the shape of their fields. Their response to a current-carrying coil and a permanent magnet is seen to be very similar, especially near the ends of the coil and the magnet.

Combining a ferromagnet with an electromagnet can produce particularly strong magnetic effects. (See [link].) Whenever strong magnetic effects are needed, such as lifting scrap metal, or in particle accelerators, electromagnets are enhanced by ferromagnetic materials. Limits to how strong the magnets can be made are imposed by coil resistance (it will overheat and melt at sufficiently high current), and so superconducting magnets may be employed. These are still limited, because superconducting properties are destroyed by too great a magnetic field.



An electromagnet with a ferromagnetic core can produce very strong magnetic effects. Alignment of domains in the core produces a magnet, the poles of which are aligned with the electromagnet.

[link] shows a few uses of combinations of electromagnets and ferromagnets. Ferromagnetic materials can act as memory devices, because the orientation of the magnetic fields of small domains can be reversed or erased. Magnetic information storage on videotapes and computer hard drives are among the most common applications. This property is vital in our digital world.


Rotating magnetic disk
Read/write head
Voice coil positioner
Recording head
Iron core
I
Ferromagnetic coating
Motion of tape
Plastic backing
Magnetic fringe field magnetizes coating on tape
S
N
N S
N S
N S
N S
N S
Induced magnetism

An electromagnet induces regions of permanent magnetism on a floppy disk coated with a ferromagnetic material. The information stored here is digital (a region is either magnetic or not); in other applications, it can be analog

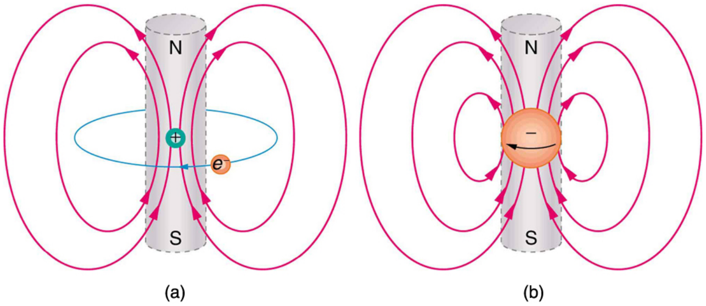
(with a varying strength), such
as on audiotapes.

Current: The Source of All Magnetism

An electromagnet creates magnetism with an electric current. In later sections we explore this more quantitatively, finding the strength and direction of magnetic fields created by various currents. But what about ferromagnets? [link] shows models of how electric currents create magnetism at the submicroscopic level. (Note that we cannot directly observe the paths of individual electrons about atoms, and so a model or visual image, consistent with all direct observations, is made. We can directly observe the electron’s orbital angular momentum, its spin momentum, and subsequent magnetic moments, all of which are explained with electric-current-creating subatomic magnetism.) Currents, including those associated with other submicroscopic particles like protons, allow us to explain ferromagnetism and all other magnetic effects. Ferromagnetism, for example, results from an internal cooperative alignment of electron spins, possible in some materials but not in others.

Crucial to the statement that electric current is the source of all magnetism is the fact that it is impossible to separate north and south magnetic poles. (This is far different from the case of positive and negative charges, which are easily separated.) A current loop always produces a magnetic dipole—that is, a magnetic field that acts like a north pole and south pole pair. Since isolated north and south magnetic poles, called **magnetic monopoles**, are not observed, currents are used to explain all magnetic effects. If magnetic monopoles did exist, then we would have to modify this underlying connection that all magnetism is due to electrical current. There is no known reason that magnetic monopoles should not exist—they are simply never observed—and so searches at the subnuclear level continue. If they do *not* exist, we would like to find out why not. If they *do* exist, we would like to see evidence of them.

Note:
Electric Currents and Magnetism
Electric current is the source of all magnetism.



(a) In the planetary model of the atom, an electron orbits a nucleus, forming a closed-current loop and producing a magnetic field with a north pole and a south pole. (b) Electrons have spin and can be crudely pictured as rotating charge, forming a current that produces a magnetic field with a north pole and a south pole. Neither the planetary model nor the image of a spinning electron is completely consistent with modern physics. However, they do provide a useful way of understanding phenomena.

Note:
PhET Explorations: Magnets and Electromagnets
Explore the interactions between a compass and bar magnet. Discover how you can use a battery and wire to make a magnet! Can you make it a stronger magnet? Can you make the magnetic field reverse?

https://archive.cnx.org/specials/92176000-ae74-11e5-baad-cfab91c15075/magnets-and-electromagnets/#sim-bar-magnet

Section Summary

- Magnetic poles always occur in pairs of north and south—it is not possible to isolate north and south poles.
- All magnetism is created by electric current.
- Ferromagnetic materials, such as iron, are those that exhibit strong magnetic effects.
- The atoms in ferromagnetic materials act like small magnets (due to currents within the atoms) and can be aligned, usually in millimeter-sized regions called domains.
- Domains can grow and align on a larger scale, producing permanent magnets. Such a material is magnetized, or induced to be magnetic.
- Above a material’s Curie temperature, thermal agitation destroys the alignment of atoms, and ferromagnetism disappears.
- Electromagnets employ electric currents to make magnetic fields, often aided by induced fields in ferromagnetic materials.

Glossary

ferromagnetic
materials, such as iron, cobalt, nickel, and gadolinium, that exhibit strong magnetic effects

magnetized
to be turned into a magnet; to be induced to be magnetic

domains
regions within a material that behave like small bar magnets

Curie temperature
the temperature above which a ferromagnetic material cannot be magnetized

electromagnetism

the use of electrical currents to induce magnetism

electromagnet

an object that is temporarily magnetic when an electrical current is passed through it

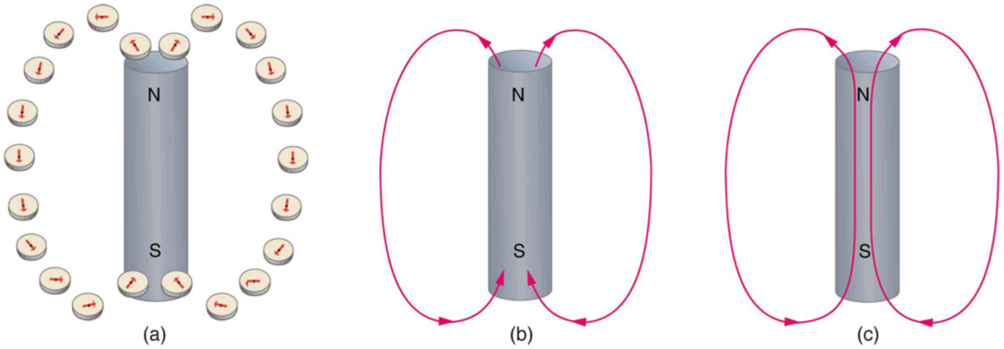
magnetic monopoles

an isolated magnetic pole; a south pole without a north pole, or vice versa (no magnetic monopole has ever been observed)

Magnetic Fields and Magnetic Field Lines

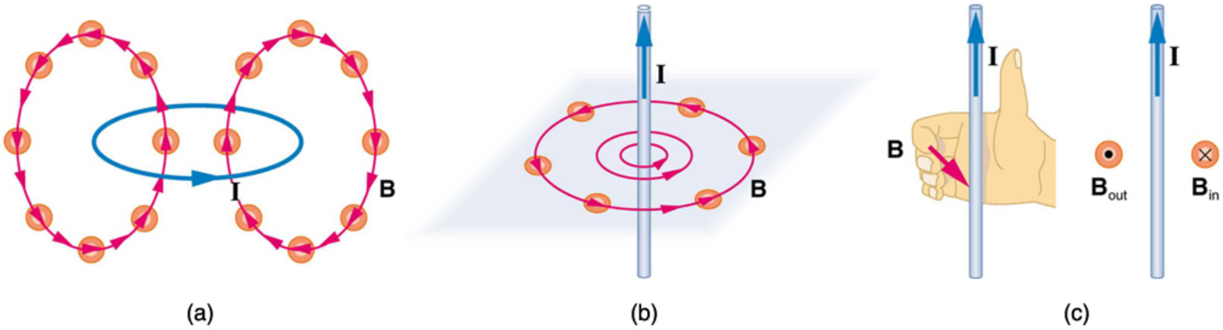
- Define magnetic field and describe the magnetic field lines of various magnetic fields.

Einstein is said to have been fascinated by a compass as a child, perhaps musing on how the needle felt a force without direct physical contact. His ability to think deeply and clearly about action at a distance, particularly for gravitational, electric, and magnetic forces, later enabled him to create his revolutionary theory of relativity. Since magnetic forces act at a distance, we define a **magnetic field** to represent magnetic forces. The pictorial representation of **magnetic field lines** is very useful in visualizing the strength and direction of the magnetic field. As shown in [link], the **direction of magnetic field lines** is defined to be the direction in which the north end of a compass needle points. The magnetic field is traditionally called the ***B*-field**.



Magnetic field lines are defined to have the direction that a small compass points when placed at a location. (a) If small compasses are used to map the magnetic field around a bar magnet, they will point in the directions shown: away from the north pole of the magnet, toward the south pole of the magnet. (Recall that the Earth's north magnetic pole is really a south pole in terms of definitions of poles on a bar magnet.) (b) Connecting the arrows gives continuous magnetic field lines. The strength of the field is proportional to the closeness (or density) of the lines. (c) If the interior of the magnet could be probed, the field lines would be found to form continuous closed loops.

Small compasses used to test a magnetic field will not disturb it. (This is analogous to the way we tested electric fields with a small test charge. In both cases, the fields represent only the object creating them and not the probe testing them.) [link] shows how the magnetic field appears for a current loop and a long straight wire, as could be explored with small compasses. A small compass placed in these fields will align itself parallel to the field line at its location, with its north pole pointing in the direction of *B*. Note the symbols used for field into and out of the paper.



Small compasses could be used to map the fields shown here. (a) The magnetic field of a circular current loop is similar to that of a bar magnet. (b) A long and straight wire creates a field with magnetic field lines forming circular loops. (c) When the wire is in the plane of the paper, the field is perpendicular to the paper. Note that the symbols used for the field pointing inward (like the tail of an arrow) and the field pointing outward (like the tip of an arrow).

Note:

Making Connections: Concept of a Field

A field is a way of mapping forces surrounding any object that can act on another object at a distance without apparent physical connection. The field represents the object generating it. Gravitational fields map gravitational forces, electric fields map electrical forces, and magnetic fields map magnetic forces.

Extensive exploration of magnetic fields has revealed a number of hard-and-fast rules. We use magnetic field lines to represent the field (the lines are a pictorial tool, not a physical entity in and of themselves). The properties of magnetic field lines can be summarized by these rules:

1. The direction of the magnetic field is tangent to the field line at any point in space. A small compass will point in the direction of the field line.
2. The strength of the field is proportional to the closeness of the lines. It is exactly proportional to the number of lines per unit area perpendicular to the lines (called the areal density).
3. Magnetic field lines can never cross, meaning that the field is unique at any point in space.
4. Magnetic field lines are continuous, forming closed loops without beginning or end. They go from the north pole to the south pole.

The last property is related to the fact that the north and south poles cannot be separated. It is a distinct difference from electric field lines, which begin and end on the positive and negative charges. If magnetic monopoles existed, then magnetic field lines would begin and end on them.

Section Summary

- Magnetic fields can be pictorially represented by magnetic field lines, the properties of which are as follows:

1. The field is tangent to the magnetic field line.
2. Field strength is proportional to the line density.
3. Field lines cannot cross.
4. Field lines are continuous loops.

Conceptual Questions

Exercise:

Problem:

Explain why the magnetic field would not be unique (that is, not have a single value) at a point in space where magnetic field lines might cross. (Consider the direction of the field at such a point.)

Exercise:

Problem:

List the ways in which magnetic field lines and electric field lines are similar. For example, the field direction is tangent to the line at any point in space. Also list the ways in which they differ. For example, electric force is parallel to electric field lines, whereas magnetic force on moving charges is perpendicular to magnetic field lines.

Exercise:

Problem:

Noting that the magnetic field lines of a bar magnet resemble the electric field lines of a pair of equal and opposite charges, do you expect the magnetic field to rapidly decrease in strength with distance from the magnet? Is this consistent with your experience with magnets?

Exercise:

Problem:

Is the Earth’s magnetic field parallel to the ground at all locations? If not, where is it parallel to the surface? Is its strength the same at all locations? If not, where is it greatest?

Glossary

magnetic field	the representation of magnetic forces

B-field
another term for magnetic field

magnetic field lines
the pictorial representation of the strength and the direction of a magnetic field

direction of magnetic field lines
the direction that the north end of a compass needle points

Magnetic Field Strength: Force on a Moving Charge in a Magnetic Field

- Describe the effects of magnetic fields on moving charges.
- Use the right hand rule 1 to determine the velocity of a charge, the direction of the magnetic field, and the direction of the magnetic force on a moving charge.
- Calculate the magnetic force on a moving charge.

What is the mechanism by which one magnet exerts a force on another? The answer is related to the fact that all magnetism is caused by current, the flow of charge. *Magnetic fields exert forces on moving charges*, and so they exert forces on other magnets, all of which have moving charges.

Right Hand Rule 1

The magnetic force on a moving charge is one of the most fundamental known. Magnetic force is as important as the electrostatic or Coulomb force. Yet the magnetic force is more complex, in both the number of factors that affects it and in its direction, than the relatively simple Coulomb force. The magnitude of the **magnetic force** F on a charge q moving at a speed v in a magnetic field of strength B is given by
Equation:

$$F=\text{qvB sin }\theta ,$$

where θ is the angle between the directions of $\mathbf{v}$ and $\mathbf{B}$. This force is often called the **Lorentz force**. In fact, this is how we define the magnetic field strength B—in terms of the force on a charged particle moving in a magnetic field. The SI unit for magnetic field strength B is called the **tesla** (T) after the eccentric but brilliant inventor Nikola Tesla (1856–1943). To determine how the tesla relates to other SI units, we solve $F=\text{qvB sin }\theta$ for B.
Equation:

$$B=\frac{F}{\text{qv sin }\theta }$$

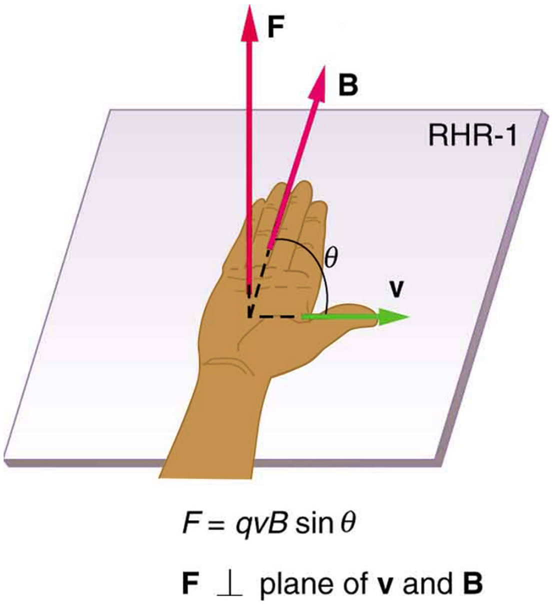
Because sin θ is unitless, the tesla is
Equation:

$$1\ \text{T}=\frac{1\ \text{N}}{\text{C}\cdot\text{m/s}}=\frac{1\ \text{N}}{\text{A}\cdot\text{m}}$$

(note that C/s = A).

Another smaller unit, called the **gauss** (G), where $1\ \text{G}=10^{-4}\ \text{T}$, is sometimes used. The strongest permanent magnets have fields near 2 T; superconducting electromagnets may attain 10 T or more. The Earth’s magnetic field on its surface is only about $5\times10^{-5}\ \text{T}$, or 0.5 G.

The *direction* of the magnetic force $\mathbf{F}$ is perpendicular to the plane formed by $\mathbf{v}$ and $\mathbf{B}$, as determined by the **right hand rule 1** (or RHR-1), which is illustrated in [link]. RHR-1 states that, to determine the direction of the magnetic force on a positive moving charge, you point the thumb of the right hand in the direction of $\mathbf{v}$, the fingers in the direction of $\mathbf{B}$, and a perpendicular to the palm points in the direction of $\mathbf{F}$. One way to remember this is that there is one velocity, and so the thumb represents it. There are many field lines, and so the fingers represent them. The force is in the direction you would push with your palm. The force on a negative charge is in exactly the opposite direction to that on a positive charge.



Magnetic fields exert forces on moving charges. This force is one of the most basic known. The direction of the magnetic force on a moving charge is perpendicular to the plane formed by $\mathbf{v}$ and $\mathbf{B}$ and follows right hand rule–1 (RHR-1) as shown. The magnitude of the force is proportional to q, v, B, and the sine of the angle between $\mathbf{v}$ and $\mathbf{B}$.

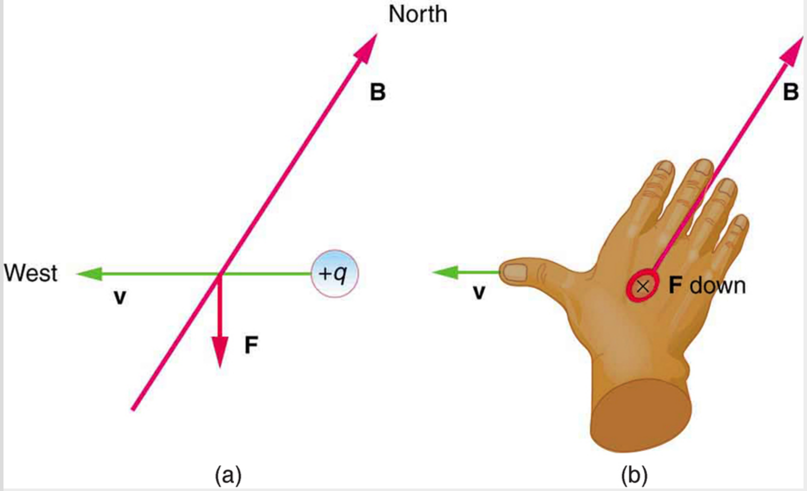
Note:

Making Connections: Charges and Magnets

There is no magnetic force on static charges. However, there is a magnetic force on moving charges. When charges are stationary, their electric fields do not affect magnets. But, when charges move, they produce magnetic fields that exert forces on other magnets. When there is relative motion, a connection between electric and magnetic fields emerges—each affects the other.

Example:
Calculating Magnetic Force: Earth’s Magnetic Field on a Charged Glass Rod

With the exception of compasses, you seldom see or personally experience forces due to the Earth’s small magnetic field. To illustrate this, suppose that in a physics lab you rub a glass rod with silk, placing a 20-nC positive charge on it. Calculate the force on the rod due to the Earth’s magnetic field, if you throw it with a horizontal velocity of 10 m/s due west in a place where the Earth’s field is due north parallel to the ground. (The direction of the force is determined with right hand rule 1 as shown in [link].)



A positively charged object moving due west in a region where the Earth’s magnetic field is due north experiences a force that is straight down as shown. A negative charge moving in the same direction would feel a force straight up.

Strategy

We are given the charge, its velocity, and the magnetic field strength and direction. We can thus use the equation $F=\text{qvB sin }\theta$ to find the force.

Solution

The magnetic force is

Equation:

$$F=\text{qvb sin }\theta.$$

We see that $\text{sin }\theta =1$, since the angle between the velocity and the direction of the field is 90º. Entering the other given quantities yields

Equation:

$$\begin{array}{lll}F& =& \left(20\times {10}^{–9}\text{ C}\right)\left(10\text{ m/s}\right)\left(5\times {10}^{–5}\text{ T}\right)\\ & =& 1\times {10}^{–11}\left(\text{C}\cdot \text{m/s}\right)\left(\frac{\text{N}}{\text{C}\cdot\text{m/s}}\right)=1\times {10}^{–11}\text{ N.}\end{array}$$

Discussion

This force is completely negligible on any macroscopic object, consistent with experience. (It is calculated to only one digit, since the Earth’s field varies with location and is given to only one digit.) The Earth’s magnetic field, however, does produce very important effects, particularly on submicroscopic particles. Some of these are explored in Force on a Moving Charge in a Magnetic Field: Examples and Applications.

Section Summary

- Magnetic fields exert a force on a moving charge q, the magnitude of which is

  Equation:

  $$F=\text{qvB sin }\theta ,$$

  where θ is the angle between the directions of v and B.
- The SI unit for magnetic field strength B is the tesla (T), which is related to other units by

Equation:

$$1\ \text{T} = \frac{1\ \text{N}}{\text{C}\cdot\text{m/s}} = \frac{1\ \text{N}}{\text{A}\cdot\text{m}}.$$

- The *direction* of the force on a moving charge is given by right hand rule 1 (RHR-1): Point the thumb of the right hand in the direction of v, the fingers in the direction of B, and a perpendicular to the palm points in the direction of F.
- The force is perpendicular to the plane formed by $\mathbf{v}$ and $\mathbf{B}$. Since the force is zero if $\mathbf{v}$ is parallel to $\mathbf{B}$, charged particles often follow magnetic field lines rather than cross them.

Conceptual Questions

Exercise:

Problem:

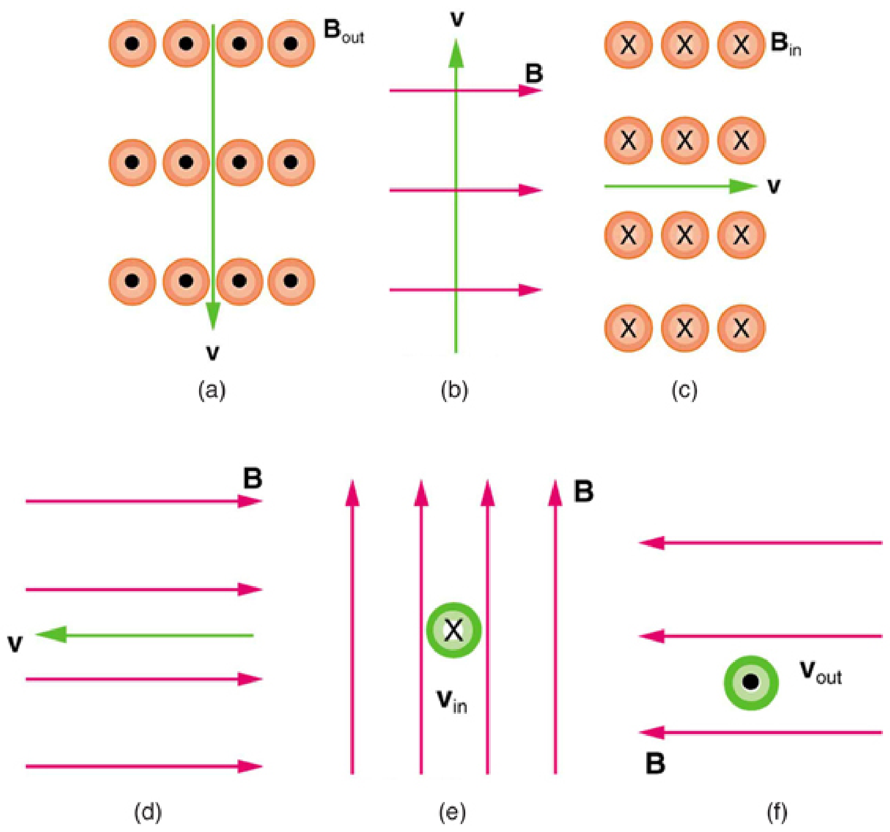
If a charged particle moves in a straight line through some region of space, can you say that the magnetic field in that region is necessarily zero?

Problems & Exercises

Exercise:

Problem:

What is the direction of the magnetic force on a positive charge that moves as shown in each of the six cases shown in [link]?


B_out
v
(a)
v
B
(b)
B_in
v
(c)
B
v
(d)
B
v_in
(e)
v_out
B
(f)

Solution:

(a) Left (West)

(b) Into the page

(c) Up (North)

(d) No force

(e) Right (East)

(f) Down (South)

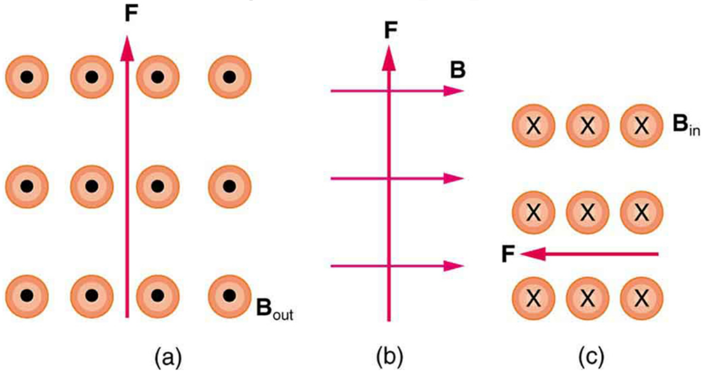
Exercise:

Problem: Repeat [link] for a negative charge.

Exercise:

Problem:

What is the direction of the velocity of a negative charge that experiences the magnetic force shown in each of the three cases in [link], assuming it moves perpendicular to $\mathbf{B}$?



Solution:

(a) East (right)

(b) Into page

(c) South (down)

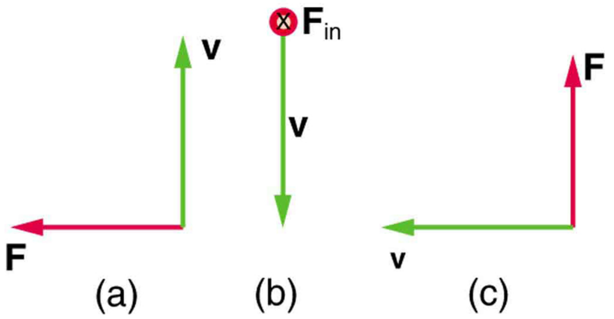
Exercise:

Problem: Repeat [link] for a positive charge.

Exercise:

Problem:

What is the direction of the magnetic field that produces the magnetic force on a positive charge as shown in each of the three cases in the figure below, assuming $\mathbf{B}$ is perpendicular to $\mathbf{v}$?



Solution:

(a) Into page

(b) West (left)

(c) Out of page

Exercise:

Problem: Repeat [link] for a negative charge.

Exercise:

Problem:

What is the maximum magnitude of the force on an aluminum rod with a $0.100\text{-}\mu\text{C}$ charge that you pass between the poles of a 1.50-T permanent magnet at a speed of 5.00 m/s? In what direction is the force?

Solution:

7.50×10^{-7} N perpendicular to both the magnetic field lines and the velocity

Exercise:

Problem:

(a) Aircraft sometimes acquire small static charges. Suppose a supersonic jet has a $0.500\text{-}\mu\text{C}$ charge and flies due west at a speed of 660 m/s over the Earth’s magnetic south pole (near Earth's geographic north pole), where the $8.00 \times 10^{-5}\text{-T}$ magnetic field points straight down. What are the direction and the magnitude of the magnetic force on the plane? (b) Discuss whether the value obtained in part (a) implies this is a significant or negligible effect.

Exercise:

Problem:

(a) A cosmic ray proton moving toward the Earth at $5.00 \times 10^{7}\ \text{m/s}$ experiences a magnetic force of $1.70 \times 10^{-16}\ \text{N}$. What is the strength of the magnetic field if there is a 45^{o} angle between it and the proton’s velocity? (b) Is the value obtained in part (a) consistent with the known strength of the Earth’s magnetic field on its surface? Discuss.

Solution:

(a) $3.01 \times 10^{-5}\ \text{T}$

(b) This is slightly less then the magnetic field strength of $5 \times 10^{-5}\ \text{T}$ at the surface of the Earth, so it is consistent.

Exercise:

Problem:

An electron moving at $4.00 \times 10^{3}\ \text{m/s}$ in a 1.25-T magnetic field experiences a magnetic force of $1.40 \times 10^{-16}\ \text{N}$. What angle does the velocity of the electron make with the magnetic field? There are two answers.

Exercise:

Problem:

(a) A physicist performing a sensitive measurement wants to limit the magnetic force on a moving charge in her equipment to less than 1.00×10^{-12} N. What is the greatest the charge can be if it moves at a maximum speed of 30.0 m/s in the Earth's field? (b) Discuss whether it would be difficult to limit the charge to less than the value found in (a) by comparing it with typical static electricity and noting that static is often absent.

Solution:

(a) 6.67×10^{-10} C (taking the Earth's field to be 5.00×10^{-5} T)

(b) Less than typical static, therefore difficult

Glossary

right hand rule 1 (RHR-1)
: the rule to determine the direction of the magnetic force on a positive moving charge: when the thumb of the right hand points in the direction of the charge's velocity $\mathbf{v}$ and the fingers point in the direction of the magnetic field $\mathbf{B}$, then the force on the charge is perpendicular and away from the palm; the force on a negative charge is perpendicular and into the palm

Lorentz force
: the force on a charge moving in a magnetic field

tesla
: T, the SI unit of the magnetic field strength; $1\ \text{T} = \frac{1\ \text{N}}{\text{A}\cdot\text{m}}$

magnetic force
: the force on a charge produced by its motion through a magnetic field; the Lorentz force

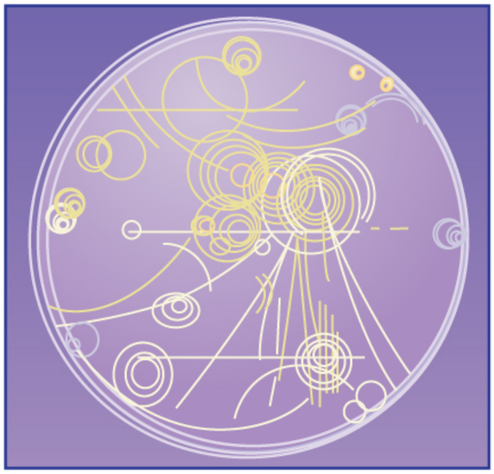
gauss

G, the unit of the magnetic field strength; $1\ \mathrm{G} = 10^{-4}\ \mathrm{T}$

Force on a Moving Charge in a Magnetic Field: Examples and Applications

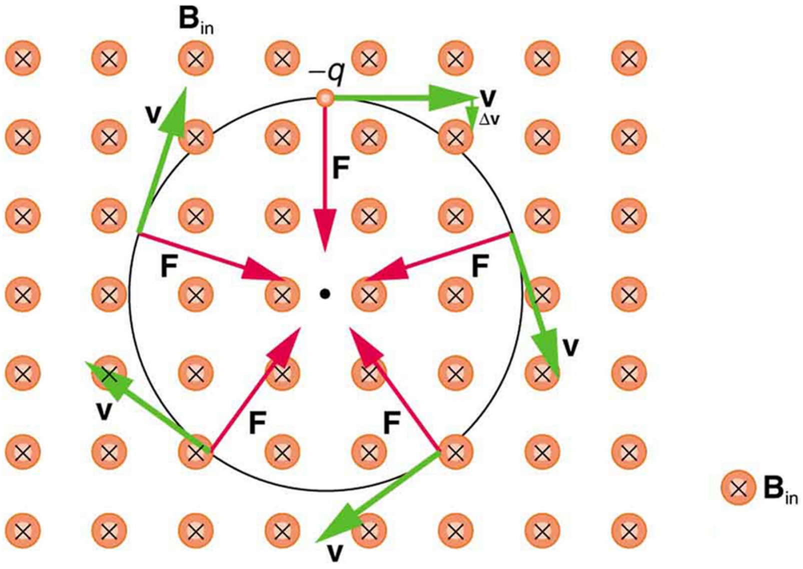
- Describe the effects of a magnetic field on a moving charge.
- Calculate the radius of curvature of the path of a charge that is moving in a magnetic field.

Magnetic force can cause a charged particle to move in a circular or spiral path. Cosmic rays are energetic charged particles in outer space, some of which approach the Earth. They can be forced into spiral paths by the Earth's magnetic field. Protons in giant accelerators are kept in a circular path by magnetic force. The bubble chamber photograph in [link] shows charged particles moving in such curved paths. The curved paths of charged particles in magnetic fields are the basis of a number of phenomena and can even be used analytically, such as in a mass spectrometer.



Trails of bubbles are produced by high-energy charged particles moving through the superheated liquid hydrogen in this artist's rendition of a bubble chamber. There is a strong magnetic field perpendicular to the page that causes the curved paths of the particles. The radius of the path can be used to find the mass, charge, and energy of the particle.

So does the magnetic force cause circular motion? Magnetic force is always perpendicular to velocity, so that it does no work on the charged particle. The particle’s kinetic energy and speed thus remain constant. The direction of motion is affected, but not the speed. This is typical of uniform circular motion. The simplest case occurs when a charged particle moves perpendicular to a uniform B-field, such as shown in [link]. (If this takes place in a vacuum, the magnetic field is the dominant factor determining the motion.) Here, the magnetic force supplies the centripetal force $F_\mathrm{c} = mv^2/r$. Noting that $\sin\theta = 1$, we see that $F = \mathrm{qvB}$.



A negatively charged particle moves in the plane of the page in a region where the magnetic field is perpendicular into the page (represented by the small circles with x’s—like the tails of arrows). The magnetic force is perpendicular to the velocity, and so velocity changes in direction but not magnitude. Uniform circular motion results.

Because the magnetic force F supplies the centripetal force F_c, we have
Equation:

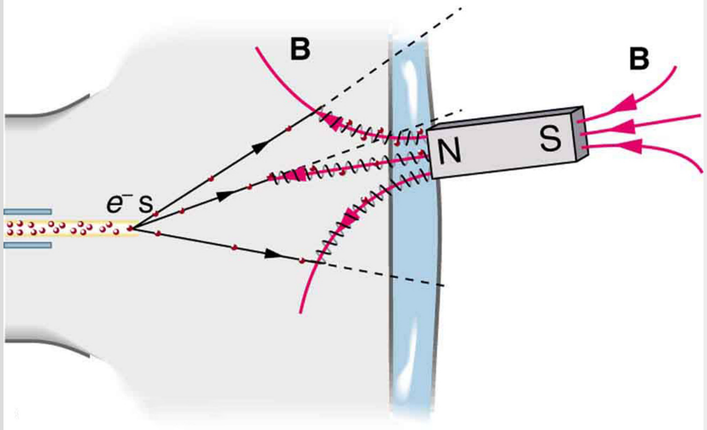
$$\text{qvB} = \frac{mv^2}{r}.$$

Solving for r yields
Equation:

$$r = \frac{\text{mv}}{\text{qB}}.$$

Here, r is the radius of curvature of the path of a charged particle with mass m and charge q, moving at a speed v perpendicular to a magnetic field of strength B. If the velocity is not perpendicular to the magnetic field, then v is the component of the velocity perpendicular to the field. The component of the velocity parallel to the field is unaffected, since the magnetic force is zero for motion parallel to the field. This produces a spiral motion rather than a circular one.

Example:
Calculating the Curvature of the Path of an Electron Moving in a Magnetic Field: A Magnet on a TV Screen
A magnet brought near an old-fashioned TV screen such as in [link] (TV sets with cathode ray tubes instead of LCD screens) severely distorts its picture by altering the path of the electrons that make its phosphors glow. ***(Don’t try this at home, as it will permanently magnetize and ruin the TV.)*** To illustrate this, calculate the radius of curvature of the path of an electron having a velocity of $6.00 \times 10^7\ \text{m/s}$ (corresponding to the accelerating voltage of about 10.0 kV used in some TVs) perpendicular to a magnetic field of strength $B = 0.500\ \text{T}$ (obtainable with permanent magnets).



Side view showing what happens when a magnet comes in contact with a computer monitor or TV screen. Electrons moving toward the screen spiral about magnetic field lines, maintaining the component of their velocity parallel to the field lines. This distorts the image on the screen.

Strategy

We can find the radius of curvature r directly from the equation $r = \frac{mv}{qB}$, since all other quantities in it are given or known.

Solution

Using known values for the mass and charge of an electron, along with the given values of v and B gives us

Equation:

$$\begin{aligned} r = \frac{\text{mv}}{\text{qB}} &= \frac{\left(9.11\times10^{-31}\ \text{kg}\right)\left(6.00\times10^{7}\ \text{m/s}\right)}{\left(1.60\times10^{-19}\ \text{C}\right)(0.500\ \text{T})} \\ &= 6.83 \times 10^{-4}\ \text{m} \end{aligned}$$

or

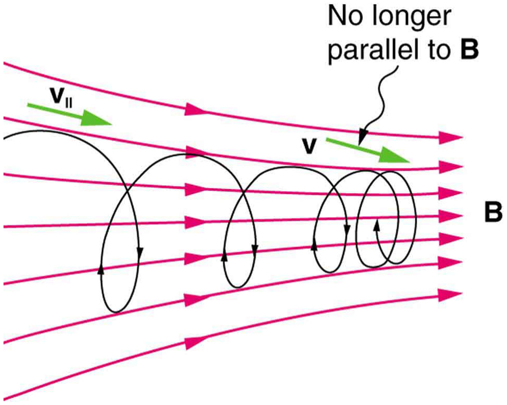
Equation:

$$r = 0.683\ \text{mm}.$$

Discussion

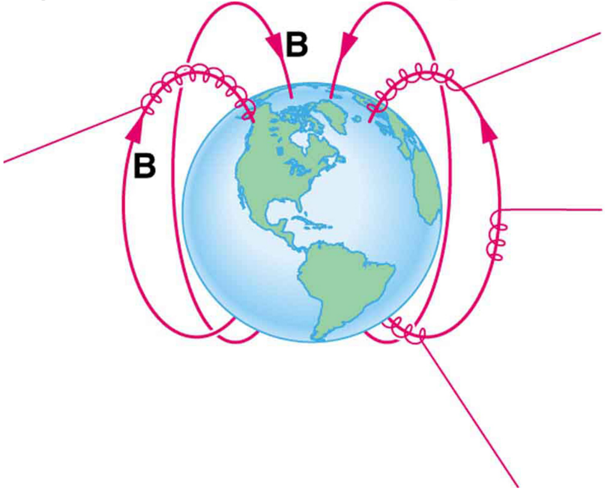
The small radius indicates a large effect. The electrons in the TV picture tube are made to move in very tight circles, greatly altering their paths and distorting the image.

[link] shows how electrons not moving perpendicular to magnetic field lines follow the field lines. The component of velocity parallel to the lines is unaffected, and so the charges spiral along the field lines. If field strength increases in the direction of motion, the field will exert a force to slow the charges, forming a kind of magnetic mirror, as shown below.



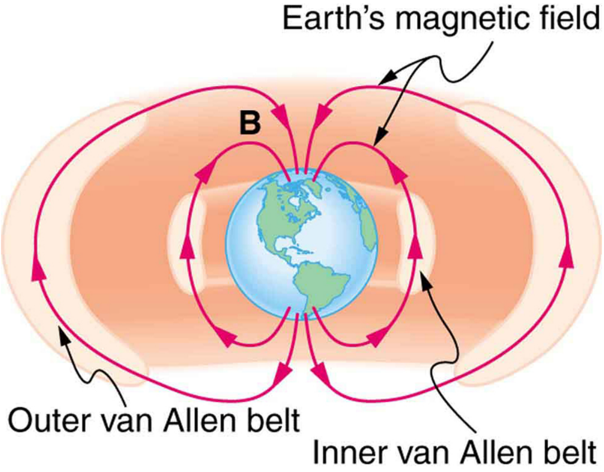
When a charged particle moves along a magnetic field line into a region where the field becomes stronger, the particle experiences a force that reduces the component of velocity parallel to the field. This force slows the motion along the field line and here reverses it, forming a “magnetic mirror.”

The properties of charged particles in magnetic fields are related to such different things as the Aurora Australis or Aurora Borealis and particle accelerators. *Charged particles approaching magnetic field lines may get trapped in spiral orbits about the lines rather than crossing them*, as seen above. Some cosmic rays, for example, follow the Earth's magnetic field lines, entering the atmosphere near the magnetic poles and causing the southern or northern lights through their ionization of molecules in the atmosphere. This glow of energized atoms and molecules is seen in [link]. Those particles that approach middle latitudes must cross magnetic field lines, and many are prevented from penetrating the atmosphere. Cosmic rays are a component of background radiation; consequently, they give a higher radiation dose at the poles than at the equator.



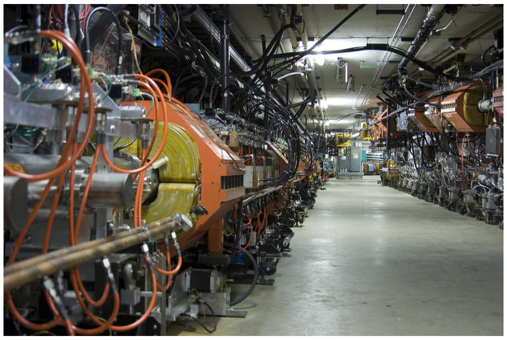
Energetic electrons and protons, components of cosmic rays, from the Sun and deep outer space often follow the Earth's magnetic field lines rather than cross them. (Recall that the Earth's north magnetic pole is really a south pole in terms of a bar magnet.)

Some incoming charged particles become trapped in the Earth’s magnetic field, forming two belts above the atmosphere known as the Van Allen radiation belts after the discoverer James A. Van Allen, an American astrophysicist. (See [link].) Particles trapped in these belts form radiation fields (similar to nuclear radiation) so intense that manned space flights avoid them and satellites with sensitive electronics are kept out of them. In the few minutes it took lunar missions to cross the Van Allen radiation belts, astronauts received radiation doses more than twice the allowed annual exposure for radiation workers. Other planets have similar belts, especially those having strong magnetic fields like Jupiter.



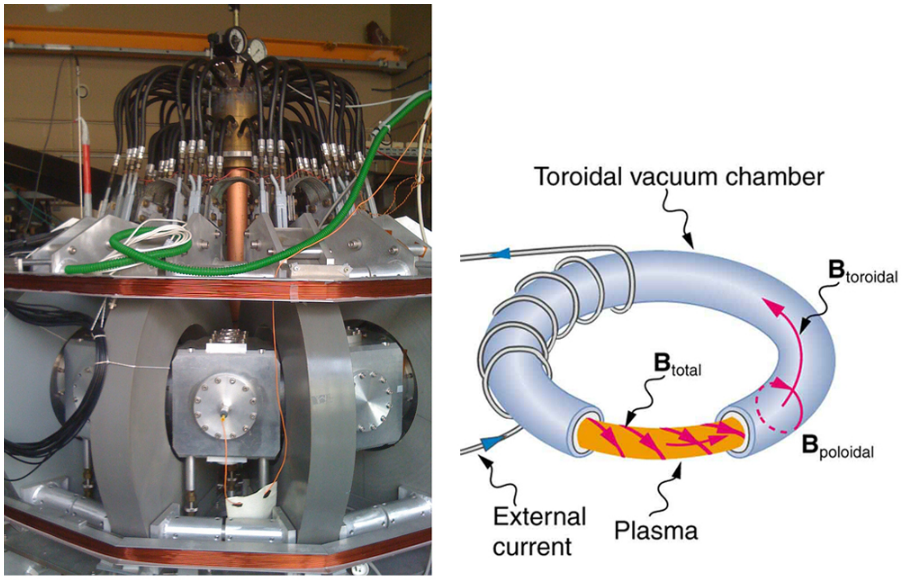
The Van Allen radiation belts are two regions in which energetic charged particles are trapped in the Earth’s magnetic field. One belt lies about 300 km above the Earth’s surface, the other about 16,000 km. Charged particles in these belts migrate along magnetic field lines and are partially reflected away from the poles by the stronger fields there. The charged particles that enter the atmosphere are replenished by the Sun and sources in deep outer space.

Back on Earth, we have devices that employ magnetic fields to contain charged particles. Among them are the giant particle accelerators that have been used to explore the substructure of matter. (See [link].) Magnetic fields not only control the direction of the charged particles, they also are used to focus particles into beams and overcome the repulsion of like charges in these beams.



The Fermilab facility in Illinois has a large particle accelerator (the most powerful in the world until 2008) that employs magnetic fields (magnets seen here in orange) to contain and direct its beam. This and other accelerators have been in use for several decades and have allowed us to discover some of the laws underlying all matter. (credit: ammcrim, Flickr)

Thermonuclear fusion (like that occurring in the Sun) is a hope for a future clean energy source. One of the most promising devices is the *tokamak*, which uses magnetic fields to contain (or trap) and direct the reactive charged particles. (See [link].) Less exotic, but more immediately practical, amplifiers in microwave ovens use a magnetic field to contain oscillating electrons. These oscillating electrons generate the microwaves sent into the oven.



(a) (b)

Tokamaks such as the one shown in the figure are being studied with the goal of economical production of energy by nuclear fusion. Magnetic fields in the doughnut-shaped device contain and direct the reactive charged particles. (credit: David Mellis, Flickr)

Mass spectrometers have a variety of designs, and many use magnetic fields to measure mass. The curvature of a charged particle’s path in the field is related to its mass and is measured to obtain mass information. (See More Applications of Magnetism.) Historically, such techniques were employed in the first direct observations of electron charge and mass. Today, mass spectrometers (sometimes coupled with gas chromatographs) are used to determine the make-up and sequencing of large biological molecules.

Section Summary

- Magnetic force can supply centripetal force and cause a charged particle to move in a circular path of radius
  Equation:
  $$r=\frac{\text{mv}}{\text{qB}},$$
  where v is the component of the velocity perpendicular to B for a charged particle with mass m and charge q.

Conceptual Questions

Exercise:

Problem:

How can the motion of a charged particle be used to distinguish between a magnetic and an electric field?

Exercise:

Problem:

High-velocity charged particles can damage biological cells and are a component of radiation exposure in a variety of locations ranging from research facilities to natural background. Describe how you could use a magnetic field to shield yourself.

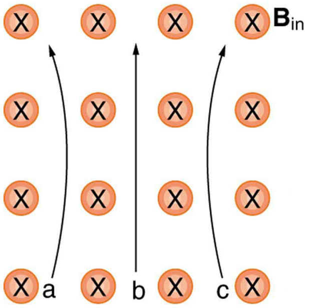
Exercise:

Problem:

If a cosmic ray proton approaches the Earth from outer space along a line toward the center of the Earth that lies in the plane of the equator, in what direction will it be deflected by the Earth’s magnetic field? What about an electron? A neutron?

Exercise:

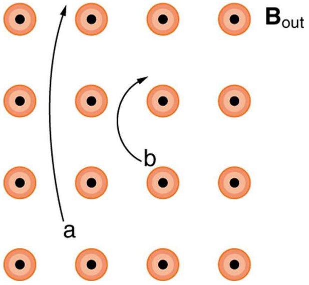
Problem: What are the signs of the charges on the particles in [link]?



Exercise:

Problem:

Which of the particles in [link] has the greatest velocity, assuming they have identical charges and masses?



Exercise:

Problem:

Which of the particles in [link] has the greatest mass, assuming all have identical charges and velocities?

Exercise:

Problem:

While operating, a high-precision TV monitor is placed on its side during maintenance. The image on the monitor changes color and blurs slightly. Discuss the possible relation of these effects to the Earth’s magnetic field.

Problems & Exercises

If you need additional support for these problems, see More Applications of Magnetism.

Exercise:

Problem:

A cosmic ray electron moves at $7.50 \times 10^{6}\ \text{m/s}$ perpendicular to the Earth’s magnetic field at an altitude where field strength is $1.00 \times 10^{-5}\ \text{T}$. What is the radius of the circular path the electron follows?

Solution:

4.27 m

Exercise:

Problem:

A proton moves at $7.50 \times 10^{7}\ \text{m/s}$ perpendicular to a magnetic field. The field causes the proton to travel in a circular path of radius 0.800 m. What is the field strength?

Exercise:

Problem:

(a) Viewers of *Star Trek* hear of an antimatter drive on the Starship *Enterprise*. One possibility for such a futuristic energy source is to store antimatter charged particles in a vacuum chamber, circulating in a magnetic field, and then extract them as needed. Antimatter annihilates with normal matter, producing pure energy. What strength magnetic field is needed to hold antiprotons, moving at $5.00 \times 10^{7}\ \text{m/s}$ in a circular path 2.00 m in radius? Antiprotons have the same mass as protons but the opposite (negative) charge. (b) Is this field strength obtainable with today's technology or is it a futuristic possibility?

Solution:

(a) 0.261 T

(b) This strength is definitely obtainable with today's technology. Magnetic field strengths of 0.500 T are obtainable with permanent magnets.

Exercise:

Problem:

(a) An oxygen-16 ion with a mass of $2.66 \times 10^{-26}\ \text{kg}$ travels at $5.00 \times 10^{6}\ \text{m/s}$ perpendicular to a 1.20-T magnetic field, which makes it move in a circular arc with a 0.231-m radius. What positive charge is on the ion? (b) What is the ratio of this charge to the charge of an electron? (c) Discuss why the ratio found in (b) should be an integer.

Exercise:

Problem:

What radius circular path does an electron travel if it moves at the same speed and in the same magnetic field as the proton in [link]?

Solution:

$4.36 \times 10^{-4}\ \text{m}$

Exercise:

Problem:

A velocity selector in a mass spectrometer uses a 0.100-T magnetic field. (a) What electric field strength is needed to select a speed of $4.00 \times 10^{6}\ \text{m/s}$? (b) What is the voltage between the plates if they are separated by 1.00 cm?

Exercise:

Problem:

An electron in a TV CRT moves with a speed of $6.00 \times 10^{7}\ \text{m/s}$, in a direction perpendicular to the Earth’s field, which has a strength of $5.00 \times 10^{-5}\ \text{T}$. (a) What strength electric field must be applied perpendicular to the Earth’s field to make the electron moves in a straight line? (b) If this is done between plates separated by 1.00 cm, what is the voltage applied? (Note that TVs are usually surrounded by a ferromagnetic material to shield against external magnetic fields and avoid the need for such a correction.)

Solution:

(a) 3.00 kV/m

(b) 30.0 V

Exercise:

Problem:

(a) At what speed will a proton move in a circular path of the same radius as the electron in [link]? (b) What would the radius of the path be if the proton had the same speed as the electron? (c) What would the radius be if the proton had the same kinetic energy as the electron? (d) The same momentum?

Exercise:

Problem:

A mass spectrometer is being used to separate common oxygen-16 from the much rarer oxygen-18, taken from a sample of old glacial ice. (The relative abundance of these oxygen isotopes is related to climatic temperature at the time the ice was deposited.) The ratio of the masses of these two ions is 16 to 18, the mass of oxygen-16 is $2.66 \times 10^{-26}\ \text{kg}$, and they are singly charged and travel at $5.00 \times 10^{6}\ \text{m/s}$ in a 1.20-T magnetic field. What is the separation between their paths when they hit a target after traversing a semicircle?

Solution:

0.173 m

Exercise:

Problem:

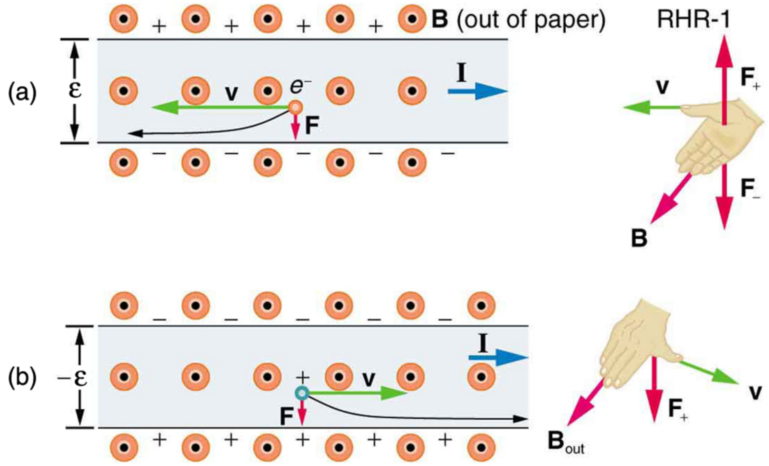
(a) Triply charged uranium-235 and uranium-238 ions are being separated in a mass spectrometer. (The much rarer uranium-235 is used as reactor fuel.) The masses of the ions are $3.90 \times 10^{-25}\ \text{kg}$ and $3.95 \times 10^{-25}\ \text{kg}$, respectively, and they travel at $3.00 \times 10^{5}\ \text{m/s}$ in a 0.250-T field. What is the separation between their paths when they hit a target after traversing a semicircle? (b) Discuss whether this distance between their paths seems to be big enough to be practical in the separation of uranium-235 from uranium-238.

The Hall Effect

- Describe the Hall effect.
- Calculate the Hall emf across a current-carrying conductor.

We have seen effects of a magnetic field on free-moving charges. The magnetic field also affects charges moving in a conductor. One result is the Hall effect, which has important implications and applications.

[link] shows what happens to charges moving through a conductor in a magnetic field. The field is perpendicular to the electron drift velocity and to the width of the conductor. Note that conventional current is to the right in both parts of the figure. In part (a), electrons carry the current and move to the left. In part (b), positive charges carry the current and move to the right. Moving electrons feel a magnetic force toward one side of the conductor, leaving a net positive charge on the other side. This separation of charge *creates a voltage* ε, known as the **Hall emf**, *across* the conductor. The creation of a voltage *across* a current-carrying conductor by a magnetic field is known as the **Hall effect**, after Edwin Hall, the American physicist who discovered it in 1879.



The Hall effect. (a) Electrons move to the left in this flat conductor (conventional current to the right). The magnetic field is directly out of the page, represented by circled dots; it exerts a force on the moving charges, causing a voltage ε, the

Hall emf, across the conductor. (b) Positive charges moving to the right (conventional current also to the right) are moved to the side, producing a Hall emf of the opposite sign, –ε. Thus, if the direction of the field and current are known, the sign of the charge carriers can be determined from the Hall effect.

One very important use of the Hall effect is to determine whether positive or negative charges carries the current. Note that in [link](b), where positive charges carry the current, the Hall emf has the sign opposite to when negative charges carry the current. Historically, the Hall effect was used to show that electrons carry current in metals and it also shows that positive charges carry current in some semiconductors. The Hall effect is used today as a research tool to probe the movement of charges, their drift velocities and densities, and so on, in materials. In 1980, it was discovered that the Hall effect is quantized, an example of quantum behavior in a macroscopic object.

The Hall effect has other uses that range from the determination of blood flow rate to precision measurement of magnetic field strength. To examine these quantitatively, we need an expression for the Hall emf, ε, across a conductor. Consider the balance of forces on a moving charge in a situation where B, v, and l are mutually perpendicular, such as shown in [link]. Although the magnetic force moves negative charges to one side, they cannot build up without limit. The electric field caused by their separation opposes the magnetic force, $F = \text{qvB}$, and the electric force, $F_{\text{e}} = \text{qE}$, eventually grows to equal it. That is,

Equation:

$$\text{qE} = \text{qvB}$$

or

Equation:

$$E = \mathrm{vB}.$$

Note that the electric field E is uniform across the conductor because the magnetic field B is uniform, as is the conductor. For a uniform electric field, the relationship between electric field and voltage is $E = \varepsilon / l$, where l is the width of the conductor and ε is the Hall emf. Entering this into the last expression gives

Equation:

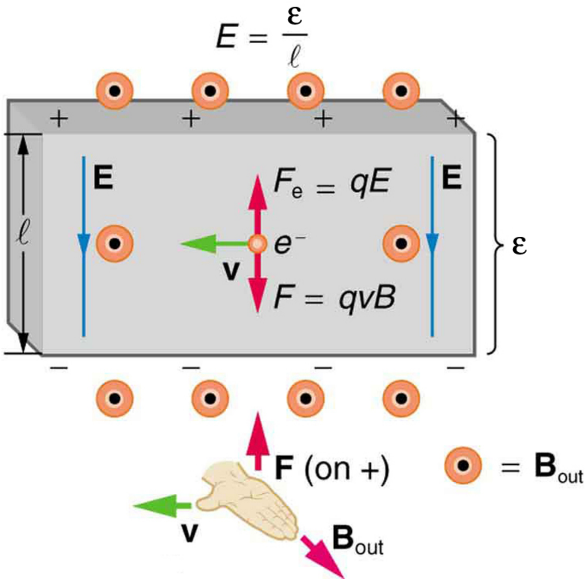
$$\frac{\varepsilon}{l} = \mathrm{vB}.$$

Solving this for the Hall emf yields

Equation:

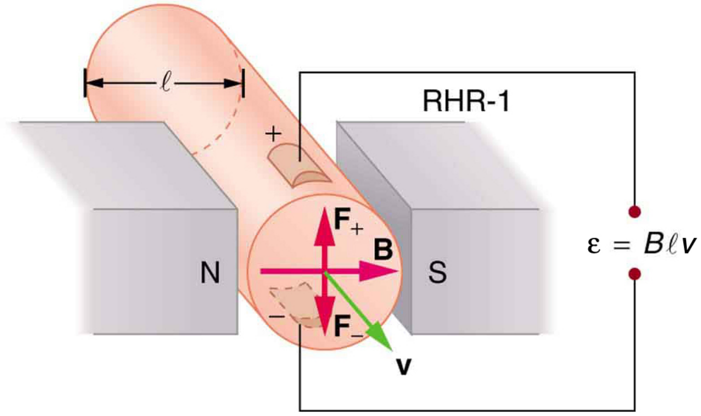
$$\varepsilon = Blv \; (B,\, v,\, \text{and } l,\, \text{mutually perpendicular}),$$

where ε is the Hall effect voltage across a conductor of width l through which charges move at a speed v.


$E = \frac{\varepsilon}{\ell}$
+ + + +
E
$F_e = qE$
E
ℓ
e^-
v
$F = qvB$
ε
− − − −
F (on +)
= $\mathbf{B}_{\text{out}}$
v
$\mathbf{B}_{\text{out}}$

The Hall emf ε produces an electric force that balances the magnetic force on the moving charges. The magnetic force produces charge separation, which builds up until it is balanced by the electric force, an equilibrium that is quickly reached.

One of the most common uses of the Hall effect is in the measurement of magnetic field strength B. Such devices, called *Hall probes*, can be made very small, allowing fine position mapping. Hall probes can also be made very accurate, usually accomplished by careful calibration. Another application of the Hall effect is to measure fluid flow in any fluid that has free charges (most do). (See [link].) A magnetic field applied perpendicular to the flow direction produces a Hall emf ε as shown. Note that the sign of ε depends not on the sign of the charges, but only on the directions of B and v. The magnitude of the Hall emf is $\varepsilon = \mathrm{Blv}$, where l is the pipe diameter, so that the average velocity v can be determined from ε providing the other factors are known.



The Hall effect can be used to measure fluid flow in any fluid having free charges, such as blood. The Hall emf ε is measured across the tube perpendicular to the applied magnetic field and is proportional to the average velocity v.

Example:

Calculating the Hall emf: Hall Effect for Blood Flow

A Hall effect flow probe is placed on an artery, applying a 0.100-T magnetic field across it, in a setup similar to that in [link]. What is the Hall emf, given the vessel's inside diameter is 4.00 mm and the average blood velocity is 20.0 cm/s?

Strategy

Because B, v, and l are mutually perpendicular, the equation $\varepsilon = \text{Blv}$ can be used to find ε.

Solution

Entering the given values for B, v, and l gives

Equation:

$$\begin{aligned} \varepsilon &= Blv = (0.100\text{ T})\left(4.00\times 10^{-3}\text{ m}\right)(0.200\text{ m/s}) \\ &= 80.0\ \mu\text{V} \end{aligned}$$

Discussion

This is the average voltage output. Instantaneous voltage varies with pulsating blood flow. The voltage is small in this type of measurement. ε is particularly difficult to measure, because there are voltages associated with heart action (ECG voltages) that are on the order of millivolts. In practice, this difficulty is overcome by applying an AC magnetic field, so that the Hall emf is AC with the same frequency. An amplifier can be very selective in picking out only the appropriate frequency, eliminating signals and noise at other frequencies.

Section Summary

- The Hall effect is the creation of voltage ε, known as the Hall emf, across a current-carrying conductor by a magnetic field.
- The Hall emf is given by

Equation:

$$\varepsilon = Blv \,(B,\, v, \text{ and } l, \text{ mutually perpendicular})$$

for a conductor of width l through which charges move at a speed v.

Conceptual Questions

Exercise:

Problem:

Discuss how the Hall effect could be used to obtain information on free charge density in a conductor. (Hint: Consider how drift velocity and current are related.)

Problems & Exercises

Exercise:

Problem:

A large water main is 2.50 m in diameter and the average water velocity is 6.00 m/s. Find the Hall voltage produced if the pipe runs perpendicular to the Earth's $5.00 \times 10^{-5}\text{-T}$ field.

Solution:

$7.50 \times 10^{-4}\text{ V}$

Exercise:

Problem:

What Hall voltage is produced by a 0.200-T field applied across a 2.60-cm-diameter aorta when blood velocity is 60.0 cm/s?

Exercise:

Problem:

(a) What is the speed of a supersonic aircraft with a 17.0-m wingspan, if it experiences a 1.60-V Hall voltage between its wing tips when in level flight over the north magnetic pole, where the Earth's field strength is $8.00 \times 10^{-5}\ \text{T}$? (b) Explain why very little current flows as a result of this Hall voltage.

Solution:

(a) 1.18×10^{3} m/s

(b) Once established, the Hall emf pushes charges one direction and the magnetic force acts in the opposite direction resulting in no net force on the charges. Therefore, no current flows in the direction of the Hall emf. This is the same as in a current-carrying conductor—current does not flow in the direction of the Hall emf.

Exercise:

Problem:

A nonmechanical water meter could utilize the Hall effect by applying a magnetic field across a metal pipe and measuring the Hall voltage produced. What is the average fluid velocity in a 3.00-cm-diameter pipe, if a 0.500-T field across it creates a 60.0-mV Hall voltage?

Exercise:

Problem:

Calculate the Hall voltage induced on a patient's heart while being scanned by an MRI unit. Approximate the conducting path on the heart wall by a wire 7.50 cm long that moves at 10.0 cm/s perpendicular to a 1.50-T magnetic field.

Solution:

11.3 mV

Exercise:

Problem:

A Hall probe calibrated to read 1.00 μV when placed in a 2.00-T field is placed in a 0.150-T field. What is its output voltage?

Exercise:

Problem:

Using information in [link], what would the Hall voltage be if a 2.00-T field is applied across a 10-gauge copper wire (2.588 mm in diameter) carrying a 20.0-A current?

Solution:

1.16 μV

Exercise:

Problem:

Show that the Hall voltage across wires made of the same material, carrying identical currents, and subjected to the same magnetic field is inversely proportional to their diameters. (Hint: Consider how drift velocity depends on wire diameter.)

Exercise:

Problem:

A patient with a pacemaker is mistakenly being scanned for an MRI image. A 10.0-cm-long section of pacemaker wire moves at a speed of 10.0 cm/s perpendicular to the MRI unit’s magnetic field and a 20.0-mV Hall voltage is induced. What is the magnetic field strength?

Solution:

2.00 T

Glossary

Hall effect

the creation of voltage across a current-carrying conductor by a magnetic field

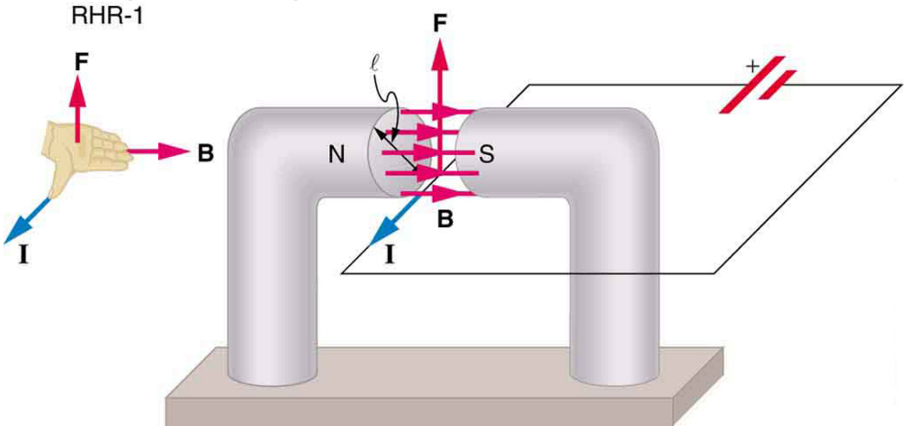
Hall emf

the electromotive force created by a current-carrying conductor by a magnetic field, $\varepsilon = \mathrm{Blv}$

Magnetic Force on a Current-Carrying Conductor

- Describe the effects of a magnetic force on a current-carrying conductor.
- Calculate the magnetic force on a current-carrying conductor.

Because charges ordinarily cannot escape a conductor, the magnetic force on charges moving in a conductor is transmitted to the conductor itself.



The magnetic field exerts a force on a current-carrying wire in a direction given by the right hand rule 1 (the same direction as that on the individual moving charges). This force can easily be large enough to move the wire, since typical currents consist of very large numbers of moving charges.

We can derive an expression for the magnetic force on a current by taking a sum of the magnetic forces on individual charges. (The forces add because they are in the same direction.) The force on an individual charge moving at the drift velocity v_{d} is given by $F = qv_{\text{d}}B\sin\theta$. Taking B to be uniform over a length of wire l and zero elsewhere, the total magnetic force on the wire is then $F = (qv_{\text{d}}B\sin\theta)(N)$, where N is the number of charge carriers in the section of wire of length l. Now, $N = \text{nV}$, where n is the number of charge carriers per unit volume and V is the volume of wire in the field. Noting that $V = \text{Al}$, where A is the cross-sectional area of the wire, then the force on the wire is $F=(qv_{\mathrm{d}}B\sin\theta)(\mathrm{nAl})$. Gathering terms,

Equation:

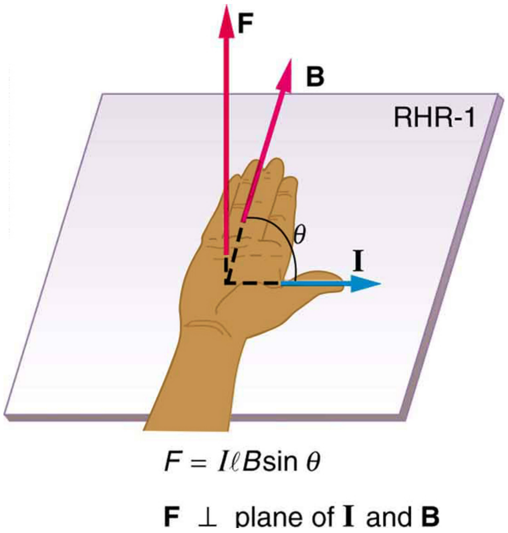
$$F=(nqAv_{\mathrm{d}})lB\sin\theta.$$

Because $nqAv_{\mathrm{d}}=I$ (see Current),

Equation:

$$F=\mathrm{IlB}\sin\theta$$

is the equation for *magnetic force on a length l of wire carrying a current I in a uniform magnetic field B*, as shown in [link]. If we divide both sides of this expression by l, we find that the magnetic force per unit length of wire in a uniform field is $\frac{F}{l}=\mathrm{IB}\sin\theta$. The direction of this force is given by RHR-1, with the thumb in the direction of the current I. Then, with the fingers in the direction of B, a perpendicular to the palm points in the direction of F, as in [link].



The force on a current-carrying wire in a magnetic field is $F=\mathrm{IlB}\sin\theta$. Its

direction is given by
RHR-1.

Example:

Calculating Magnetic Force on a Current-Carrying Wire: A Strong Magnetic Field

Calculate the force on the wire shown in [link], given $B=1.50\text{ T}$, $l=5.00\text{ cm}$, and $I=20.0\text{ A}$.

Strategy

The force can be found with the given information by using $F=\text{IlB sin }\theta$ and noting that the angle θ between I and B is 90º, so that $\text{sin }\theta =1$.

Solution

Entering the given values into $F=\text{IlB sin }\theta$ yields

Equation:

$$F=\text{IlB sin }\theta =\left(\text{20.0 A}\right)\left(\text{0.0500 m}\right)\left(\text{1.50 T}\right)\left(1\right).$$

The units for tesla are $1\text{ T}=\frac{\text{N}}{\text{A}\cdot \text{m}}$; thus,

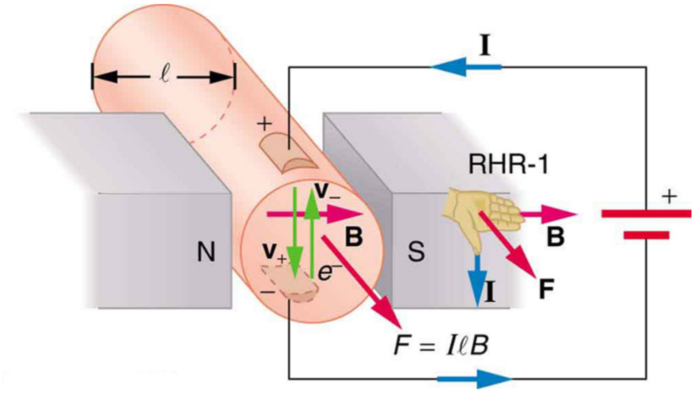
Equation:

$$F=\text{1.50 N.}$$

Discussion

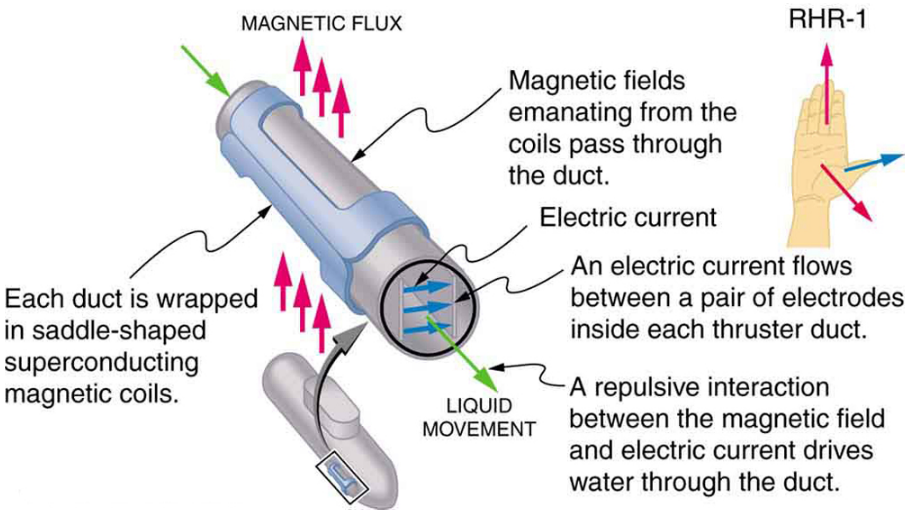
This large magnetic field creates a significant force on a small length of wire.

Magnetic force on current-carrying conductors is used to convert electric energy to work. (Motors are a prime example—they employ loops of wire and are considered in the next section.) Magnetohydrodynamics (MHD) is the technical name given to a clever application where magnetic force pumps fluids without moving mechanical parts. (See [link].)



Magnetohydrodynamics. The magnetic force on the current passed through this fluid can be used as a nonmechanical pump.

A strong magnetic field is applied across a tube and a current is passed through the fluid at right angles to the field, resulting in a force on the fluid parallel to the tube axis as shown. The absence of moving parts makes this attractive for moving a hot, chemically active substance, such as the liquid sodium employed in some nuclear reactors. Experimental artificial hearts are testing with this technique for pumping blood, perhaps circumventing the adverse effects of mechanical pumps. (Cell membranes, however, are affected by the large fields needed in MHD, delaying its practical application in humans.) MHD propulsion for nuclear submarines has been proposed, because it could be considerably quieter than conventional propeller drives. The deterrent value of nuclear submarines is based on their ability to hide and survive a first or second nuclear strike. As we slowly disassemble our nuclear weapons arsenals, the submarine branch will be the last to be decommissioned because of this ability (See [link].) Existing MHD drives are heavy and inefficient—much development work is needed.


MAGNETIC FLUX
RHR-1
Magnetic fields emanating from the coils pass through the duct.
Electric current
An electric current flows between a pair of electrodes inside each thruster duct.
Each duct is wrapped in saddle-shaped superconducting magnetic coils.
LIQUID MOVEMENT
A repulsive interaction between the magnetic field and electric current drives water through the duct.

An MHD propulsion system in a nuclear submarine could produce significantly less turbulence than propellers and allow it to run more silently. The development of a silent drive submarine was dramatized in the book and the film *The Hunt for Red October*.

Section Summary

- The magnetic force on current-carrying conductors is given by **Equation:**

  $$F = \text{IlB}\ \sin\ \theta,$$

  where I is the current, l is the length of a straight conductor in a uniform magnetic field B, and θ is the angle between I and B. The force follows RHR-1 with the thumb in the direction of I.

Conceptual Questions

Exercise:

Problem:

Draw a sketch of the situation in [link] showing the direction of electrons carrying the current, and use RHR-1 to verify the direction of the force on the wire.

Exercise:

Problem:

Verify that the direction of the force in an MHD drive, such as that in [link], does not depend on the sign of the charges carrying the current across the fluid.

Exercise:

Problem:

Why would a magnetohydrodynamic drive work better in ocean water than in fresh water? Also, why would superconducting magnets be desirable?

Exercise:

Problem:

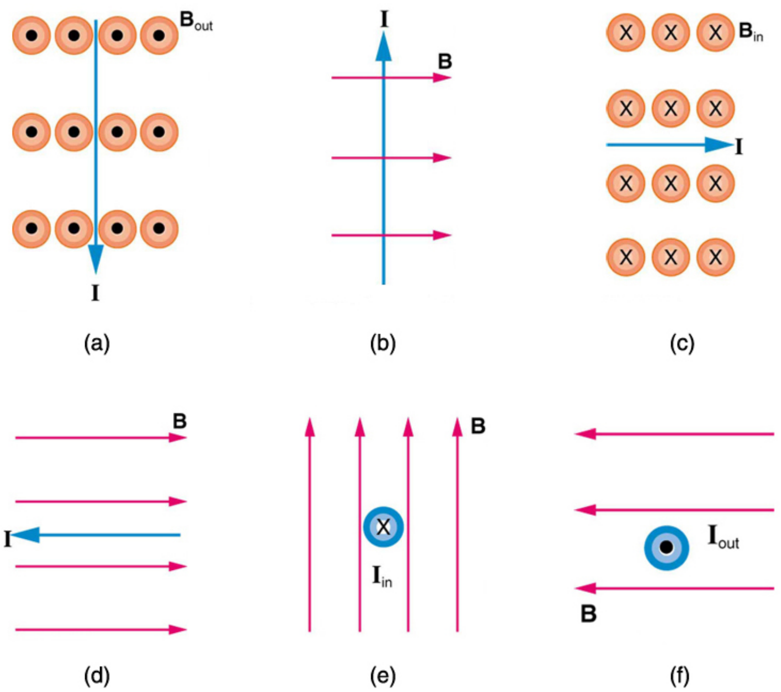
Which is more likely to interfere with compass readings, AC current in your refrigerator or DC current when you start your car? Explain.

Problems & Exercises

Exercise:

Problem:

What is the direction of the magnetic force on the current in each of the six cases in [link]?



(a) (b) (c)

(d) (e) (f)

Solution:

(a) west (left)

(b) into page

(c) north (up)

(d) no force

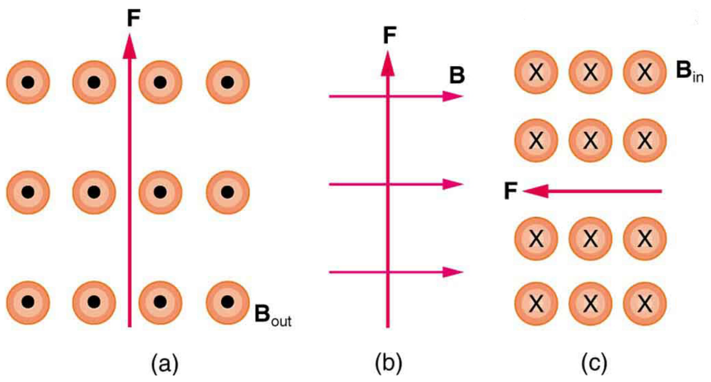
(e) east (right)

(f) south (down)

Exercise:

Problem:

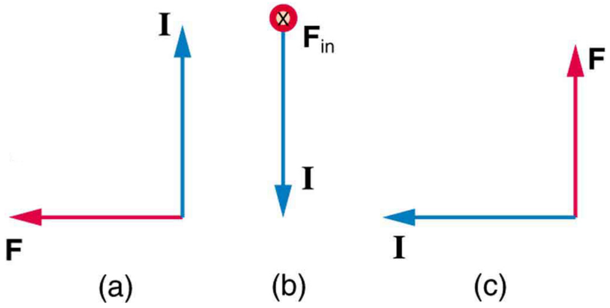
What is the direction of a current that experiences the magnetic force shown in each of the three cases in [link], assuming the current runs perpendicular to B?



Exercise:

Problem:

What is the direction of the magnetic field that produces the magnetic force shown on the currents in each of the three cases in [link], assuming $\mathbf{B}$ is perpendicular to $\mathbf{I}$?



Solution:

(a) into page

(b) west (left)

(c) out of page

Exercise:

Problem:

(a) What is the force per meter on a lightning bolt at the equator that carries 20,000 A perpendicular to the Earth’s $3.00 \times 10^{-5}\text{-T}$ field? (b) What is the direction of the force if the current is straight up and the Earth’s field direction is due north, parallel to the ground?

Exercise:

Problem:

(a) A DC power line for a light-rail system carries 1000 A at an angle of 30.0^{o} to the Earth's 5.00×10^{-5} -T field. What is the force on a 100-m section of this line? (b) Discuss practical concerns this presents, if any.

Solution:

(a) 2.50 N

(b) This is about half a pound of force per 100 m of wire, which is much less than the weight of the wire itself. Therefore, it does not cause any special concerns.

Exercise:

Problem:

What force is exerted on the water in an MHD drive utilizing a 25.0-cm-diameter tube, if 100-A current is passed across the tube that is perpendicular to a 2.00-T magnetic field? (The relatively small size of this force indicates the need for very large currents and magnetic fields to make practical MHD drives.)

Exercise:

Problem:

A wire carrying a 30.0-A current passes between the poles of a strong magnet that is perpendicular to its field and experiences a 2.16-N force on the 4.00 cm of wire in the field. What is the average field strength?

Solution:

1.80 T

Exercise:

Problem:

(a) A 0.750-m-long section of cable carrying current to a car starter motor makes an angle of 60° with the Earth's $5.50 \times 10^{-5}\ \text{T}$ field. What is the current when the wire experiences a force of $7.00 \times 10^{-3}\ \text{N}$? (b) If you run the wire between the poles of a strong horseshoe magnet, subjecting 5.00 cm of it to a 1.75-T field, what force is exerted on this segment of wire?

Exercise:

Problem:

(a) What is the angle between a wire carrying an 8.00-A current and the 1.20-T field it is in if 50.0 cm of the wire experiences a magnetic force of 2.40 N? (b) What is the force on the wire if it is rotated to make an angle of 90° with the field?

Solution:

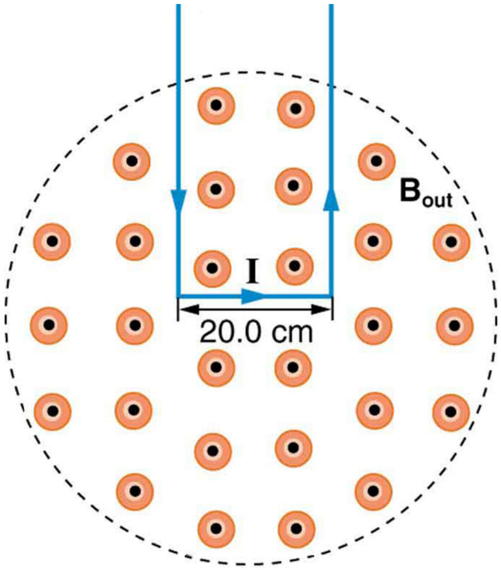
(a) 30°

(b) 4.80 N

Exercise:

Problem:

The force on the rectangular loop of wire in the magnetic field in [link] can be used to measure field strength. The field is uniform, and the plane of the loop is perpendicular to the field. (a) What is the direction of the magnetic force on the loop? Justify the claim that the forces on the sides of the loop are equal and opposite, independent of how much of the loop is in the field and do not affect the net force on the loop. (b) If a current of 5.00 A is used, what is the force per tesla on the 20.0-cm-wide loop?

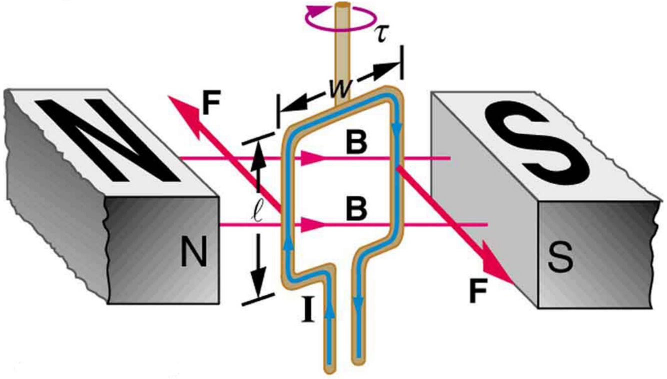


A rectangular loop of wire carrying a current is perpendicular to a magnetic field. The field is uniform in the region shown and is zero outside that region.

Torque on a Current Loop: Motors and Meters

- Describe how motors and meters work in terms of torque on a current loop.
- Calculate the torque on a current-carrying loop in a magnetic field.

Motors are the most common application of magnetic force on current-carrying wires. Motors have loops of wire in a magnetic field. When current is passed through the loops, the magnetic field exerts torque on the loops, which rotates a shaft. Electrical energy is converted to mechanical work in the process. (See [link].)

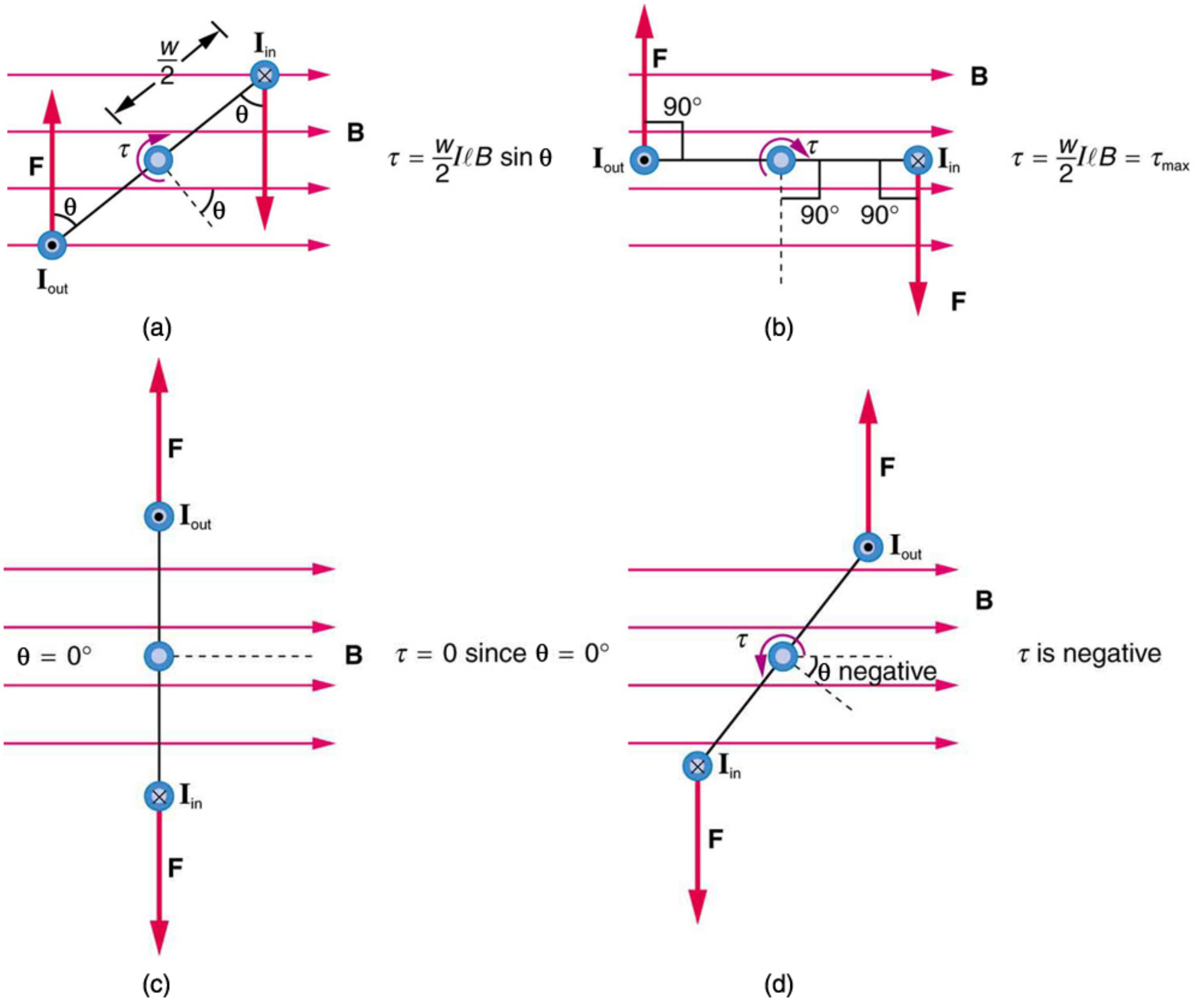


Torque on a current loop. A current-carrying loop of wire attached to a vertically rotating shaft feels magnetic forces that produce a clockwise torque as viewed from above.

Let us examine the force on each segment of the loop in [link] to find the torques produced about the axis of the vertical shaft. (This will lead to a useful equation for the torque on the loop.) We take the magnetic field to be uniform over the rectangular loop, which has width w and height l. First, we note that the forces on the top and bottom segments are vertical and, therefore, parallel to the shaft, producing no torque. Those vertical forces are equal in magnitude and opposite in direction, so that they also produce no net force on the loop. [link] shows views of the loop from above. Torque is defined as $\tau = \text{rF}\sin\theta$, where F is the force, r is the distance from the pivot that the force is applied, and θ is the angle between r and F. As seen in [link](a), right hand rule 1 gives the forces on the sides to be equal in magnitude and opposite in direction, so that the net force is again zero. However, each force produces a clockwise torque. Since $r = w/2$, the torque on each vertical segment is $(w/2)F\sin\theta$, and the two add to give a total torque.

Equation:

$$\tau = \frac{w}{2}F\sin\theta + \frac{w}{2}F\sin\theta = \text{wF}\sin\theta$$



$\frac{w}{2}$
I_{in}
θ
B
F
τ
θ
θ
I_{out}
$\tau = \frac{w}{2}I\ell B\sin\theta$
(a)
F
B
90°
I_{out}
τ
I_{in}
90°
90°
F
$\tau = \frac{w}{2}I\ell B = \tau_{max}$
(b)
F
I_{out}
θ = 0°
B
$\tau = 0$ since $\theta = 0°$
I_{in}
F
(c)
F
I_{out}
B
τ
θ negative
τ is negative
I_{in}
F
(d)

Top views of a current-carrying loop in a magnetic field. (a) The equation for torque is derived using this view. Note that the perpendicular to the loop makes an angle θ with the field that is the same as the angle between $w/2$ and $\mathbf{F}$. (b) The maximum torque occurs when θ is a right angle and $\sin\theta = 1$. (c) Zero (minimum) torque occurs when θ is zero and $\sin\theta = 0$. (d) The torque reverses once the loop rotates past $\theta = 0$.

Now, each vertical segment has a length l that is perpendicular to B, so that the force on each is $F = \text{IlB}$. Entering F into the expression for torque yields

Equation:

$$\tau = \text{wIlB}\sin\theta.$$

If we have a multiple loop of N turns, we get N times the torque of one loop. Finally, note that the area of the loop is $A = \text{wl}$; the expression for the torque becomes

Equation:

$$\tau = \text{NIAB}\sin\theta.$$

This is the torque on a current-carrying loop in a uniform magnetic field. This equation can be shown to be valid for a loop of any shape. The loop carries a current I, has N turns, each of area A, and the perpendicular to the loop makes an angle θ with the field B. The net force on the loop is zero.

Example:

Calculating Torque on a Current-Carrying Loop in a Strong Magnetic Field

Find the maximum torque on a 100-turn square loop of a wire of 10.0 cm on a side that carries 15.0 A of current in a 2.00-T field.

Strategy

Torque on the loop can be found using $\tau = \text{NIAB}\sin\theta$. Maximum torque occurs when $\theta = 90^{\text{o}}$ and $\sin\theta = 1$.

Solution

For $\sin\theta = 1$, the maximum torque is

Equation:

$$\tau_{\text{max}} = \text{NIAB}.$$

Entering known values yields

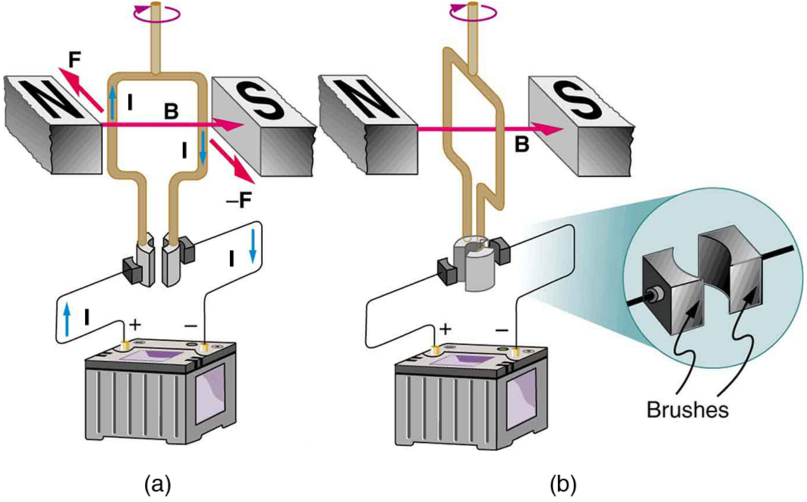
Equation:

$$\begin{aligned}\tau_{\text{max}} &= (100)(15.0\text{ A})\left(0.100\text{ m}^2\right)(2.00\text{ T})\\ &= 30.0\text{ N}\cdot\text{m}.\end{aligned}$$

Discussion

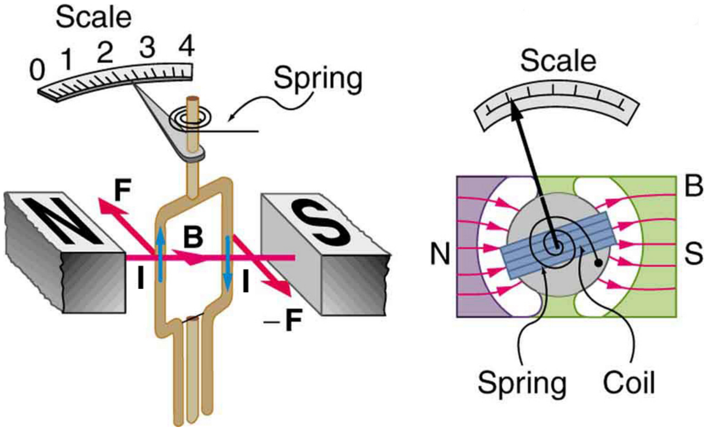
This torque is large enough to be useful in a motor.

The torque found in the preceding example is the maximum. As the coil rotates, the torque decreases to zero at $\theta = 0$. The torque then *reverses* its direction once the coil rotates past $\theta = 0$. (See [link](d).) This means that, unless we do something, the coil will oscillate back and forth about equilibrium at $\theta = 0$. To get the coil to continue rotating in the same direction, we can reverse the current as it passes through $\theta = 0$ with automatic switches called *brushes*. (See [link].)



(a) As the angular momentum of the coil carries it through $\theta = 0$, the brushes reverse the current to keep the torque clockwise. (b) The coil will rotate continuously in the clockwise direction, with the current reversing each half revolution to maintain the clockwise torque.

Meters, such as those in analog fuel gauges on a car, are another common application of magnetic torque on a current-carrying loop. [link] shows that a meter is very similar in construction to a motor. The meter in the figure has its magnets shaped to limit the effect of θ by making B perpendicular to the loop over a large angular range. Thus the torque is proportional to I and not θ. A linear spring exerts a counter-torque that balances the current-produced torque. This makes the needle deflection proportional to I. If an exact proportionality cannot be achieved, the gauge reading can be calibrated. To produce a galvanometer for use in analog voltmeters and ammeters that have a low resistance and respond to small currents, we use a large loop area A, high magnetic field B, and low-resistance coils.



Meters are very similar to motors but only rotate through a part of a revolution. The magnetic poles of this meter are shaped to keep the component of B perpendicular to the loop constant, so that the torque does not depend on θ and the deflection against the return spring is proportional only to the current I.

Section Summary

- The torque τ on a current-carrying loop of any shape in a uniform magnetic field. is
  Equation:

  $$\tau = \text{NIAB}\sin\theta,$$

  where N is the number of turns, I is the current, A is the area of the loop, B is the magnetic field strength, and θ is the angle between the perpendicular to the loop and the magnetic field.

Conceptual Questions

Exercise:

Problem:

Draw a diagram and use RHR-1 to show that the forces on the top and bottom segments of the motor's current loop in [link] are vertical and produce no torque about the axis of rotation.

Problems & Exercises

Exercise:

Problem:

(a) By how many percent is the torque of a motor decreased if its permanent magnets lose 5.0% of their strength? (b) How many percent would the current need to be increased to return the torque to original values?

Solution:

(a) τ decreases by 5.00% if B decreases by 5.00%

(b) 5.26% increase

Exercise:

Problem:

(a) What is the maximum torque on a 150-turn square loop of wire 18.0 cm on a side that carries a 50.0-A current in a 1.60-T field? (b) What is the torque when θ is 10.9º?

Exercise:

Problem:

Find the current through a loop needed to create a maximum torque of $9.00\ \mathrm{N \cdot m}$. The loop has 50 square turns that are 15.0 cm on a side and is in a uniform 0.800-T magnetic field.

Solution:

10.0 A

Exercise:

Problem:

Calculate the magnetic field strength needed on a 200-turn square loop 20.0 cm on a side to create a maximum torque of $300\ \mathrm{N \cdot m}$ if the loop is carrying 25.0 A.

Exercise:

Problem:

Since the equation for torque on a current-carrying loop is $\tau = \mathrm{NIAB}\sin\theta$, the units of $\mathrm{N \cdot m}$ must equal units of $\mathrm{A \cdot m^2\ T}$. Verify this.

Solution:

$$A \cdot m^2 \cdot T = A \cdot m^2 \left(\frac{N}{A \cdot m}\right) = N \cdot m.$$

Exercise:

Problem:

(a) At what angle θ is the torque on a current loop 90.0% of maximum? (b) 50.0% of maximum? (c) 10.0% of maximum?

Exercise:

Problem:

A proton has a magnetic field due to its spin on its axis. The field is similar to that created by a circular current loop $0.650 \times 10^{-15}\ \text{m}$ in radius with a current of $1.05 \times 10^{4}\ \text{A}$ (no kidding). Find the maximum torque on a proton in a 2.50-T field. (This is a significant torque on a small particle.)

Solution:

$3.48 \times 10^{-26}\ \text{N} \cdot \text{m}$

Exercise:

Problem:

(a) A 200-turn circular loop of radius 50.0 cm is vertical, with its axis on an east-west line. A current of 100 A circulates clockwise in the loop when viewed from the east. The Earth’s field here is due north, parallel to the ground, with a strength of $3.00 \times 10^{-5}\ \text{T}$. What are the direction and magnitude of the torque on the loop? (b) Does this device have any practical applications as a motor?

Exercise:

Problem:

Repeat [link], but with the loop lying flat on the ground with its current circulating counterclockwise (when viewed from above) in a location where the Earth’s field is north, but at an angle 45.0° below the horizontal and with a strength of $6.00 \times 10^{-5}\ \text{T}$.

Solution:

(a) $0.666\ \text{N}\cdot\text{m}$ west

(b) This is not a very significant torque, so practical use would be limited. Also, the current would need to be alternated to make the loop rotate (otherwise it would oscillate).

Glossary

motor	loop of wire in a magnetic field; when current is passed through the loops, the magnetic field exerts torque on the loops, which rotates a shaft; electrical energy is converted to mechanical work in the process
meter	common application of magnetic torque on a current-carrying loop that is very similar in construction to a motor; by design, the torque is proportional to I and not θ, so the needle deflection is proportional to the current

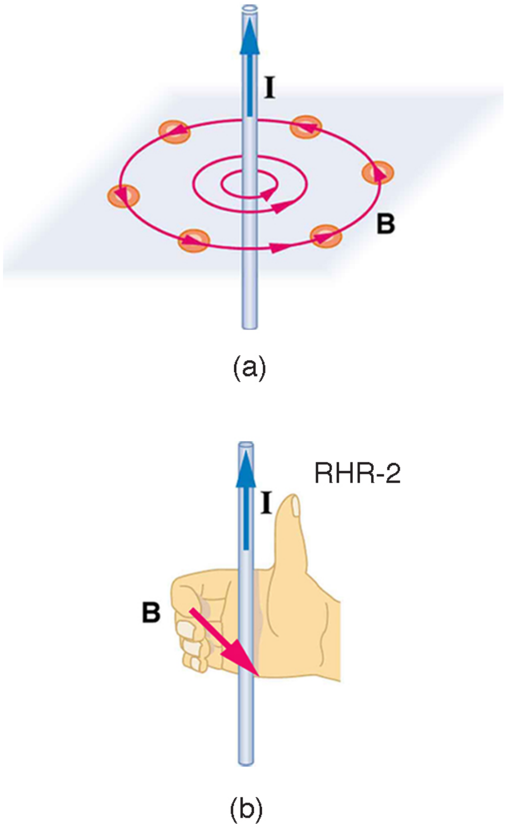
Magnetic Fields Produced by Currents: Ampere's Law

- Calculate current that produces a magnetic field.
- Use the right hand rule 2 to determine the direction of current or the direction of magnetic field loops.

How much current is needed to produce a significant magnetic field, perhaps as strong as the Earth's field? Surveyors will tell you that overhead electric power lines create magnetic fields that interfere with their compass readings. Indeed, when Oersted discovered in 1820 that a current in a wire affected a compass needle, he was not dealing with extremely large currents. How does the shape of wires carrying current affect the shape of the magnetic field created? We noted earlier that a current loop created a magnetic field similar to that of a bar magnet, but what about a straight wire or a toroid (doughnut)? How is the direction of a current-created field related to the direction of the current? Answers to these questions are explored in this section, together with a brief discussion of the law governing the fields created by currents.

Magnetic Field Created by a Long Straight Current-Carrying Wire: Right Hand Rule 2

Magnetic fields have both direction and magnitude. As noted before, one way to explore the direction of a magnetic field is with compasses, as shown for a long straight current-carrying wire in [link]. Hall probes can determine the magnitude of the field. The field around a long straight wire is found to be in circular loops. The **right hand rule 2** (RHR-2) emerges from this exploration and is valid for any current segment—*point the thumb in the direction of the current, and the fingers curl in the direction of the magnetic field loops* created by it.



(a) Compasses placed near a long straight current-carrying wire indicate that field lines form circular loops centered on the wire. (b) Right hand rule 2 states that, if the right hand thumb points in the direction of the current, the fingers curl in the direction of the field. This rule is consistent with the field mapped for the long straight wire and is valid for any current segment.

The **magnetic field strength (magnitude) produced by a long straight current-carrying wire** is found by experiment to be

Equation:

$$B = \frac{\mu_0 I}{2\pi r} \text{ (long straight wire)},$$

where I is the current, r is the shortest distance to the wire, and the constant $\mu_0 = 4\pi \times 10^{-7}\ \text{T} \cdot \text{m/A}$ is the **permeability of free space**. (μ_0 is one of the basic constants in nature. We will see later that μ_0 is related to the speed of light.) Since the wire is very long, the magnitude of the field depends only on distance from the wire r, not on position along the wire.

Example:

Calculating Current that Produces a Magnetic Field

Find the current in a long straight wire that would produce a magnetic field twice the strength of the Earth’s at a distance of 5.0 cm from the wire.

Strategy

The Earth’s field is about 5.0×10^{-5} T, and so here B due to the wire is taken to be 1.0×10^{-4} T. The equation $B = \frac{\mu_0 I}{2\pi r}$ can be used to find I, since all other quantities are known.

Solution

Solving for I and entering known values gives

Equation:

$$\begin{aligned} I &= \frac{2\pi r B}{\mu_0} = \frac{2\pi \left(5.0\times10^{-2}\text{ m}\right)\left(1.0\times10^{-4}\text{ T}\right)}{4\pi\times10^{-7}\text{ T}\cdot\text{m/A}} \\ &= 25\text{ A}. \end{aligned}$$

Discussion

So a moderately large current produces a significant magnetic field at a distance of 5.0 cm from a long straight wire. Note that the answer is stated to only two digits, since the Earth’s field is specified to only two digits in this example.

Ampere’s Law and Others

The magnetic field of a long straight wire has more implications than you might at first suspect. *Each segment of current produces a magnetic field like that of a long straight wire, and the total field of any shape current is the vector sum of the fields due to each segment.* The formal statement of the direction and magnitude of the field due to each segment is called the **Biot-Savart law**. Integral calculus is needed to sum the field for an arbitrary shape current. This results in a more complete law, called **Ampere’s law**, which relates magnetic field and current in a general way. Ampere’s law in turn is a part of **Maxwell’s equations**, which give a complete theory of all electromagnetic phenomena. Considerations of how Maxwell’s equations appear to different observers led to the modern theory of relativity, and the realization that electric and magnetic fields are different manifestations of the same thing. Most of this is beyond the scope of this text in both mathematical level, requiring calculus, and in the amount of space that can be devoted to it. But for the interested student, and particularly for those who continue in physics, engineering, or similar pursuits, delving into these matters further will reveal descriptions of nature that are elegant as well as profound. In this text, we shall keep the general features in mind, such as RHR-2 and the rules for magnetic field lines listed in Magnetic Fields and Magnetic Field Lines, while concentrating on the fields created in certain important situations.

Note:
Making Connections: Relativity
Hearing all we do about Einstein, we sometimes get the impression that he invented relativity out of nothing. On the contrary, one of Einstein’s motivations was to solve difficulties in knowing how different observers see magnetic and electric fields.

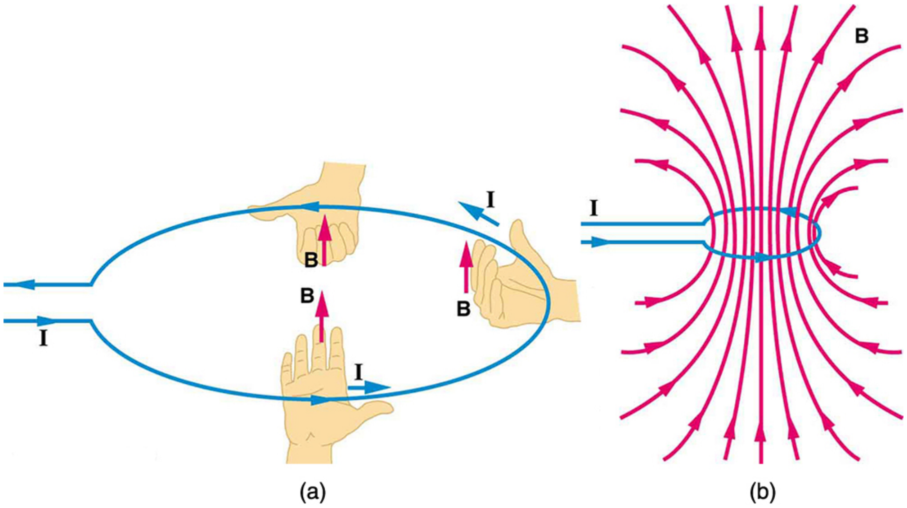
Magnetic Field Produced by a Current-Carrying Circular Loop

The magnetic field near a current-carrying loop of wire is shown in [link]. Both the direction and the magnitude of the magnetic field produced by a current-carrying loop are complex. RHR-2 can be used to give the direction of the field near the loop, but mapping with compasses and the rules about field lines given in Magnetic Fields and Magnetic Field Lines are needed for more detail. There is a simple formula for the **magnetic field strength at the center of a circular loop**. It is

Equation:

$$B = \frac{\mu_0 I}{2R}\ \text{(at center of loop)},$$

where R is the radius of the loop. This equation is very similar to that for a straight wire, but it is valid *only* at the center of a circular loop of wire. The similarity of the equations does indicate that similar field strength can be obtained at the center of a loop. One way to get a larger field is to have N loops; then, the field is $B = N\mu_0 I/(2R)$. Note that the larger the loop, the smaller the field at its center, because the current is farther away.

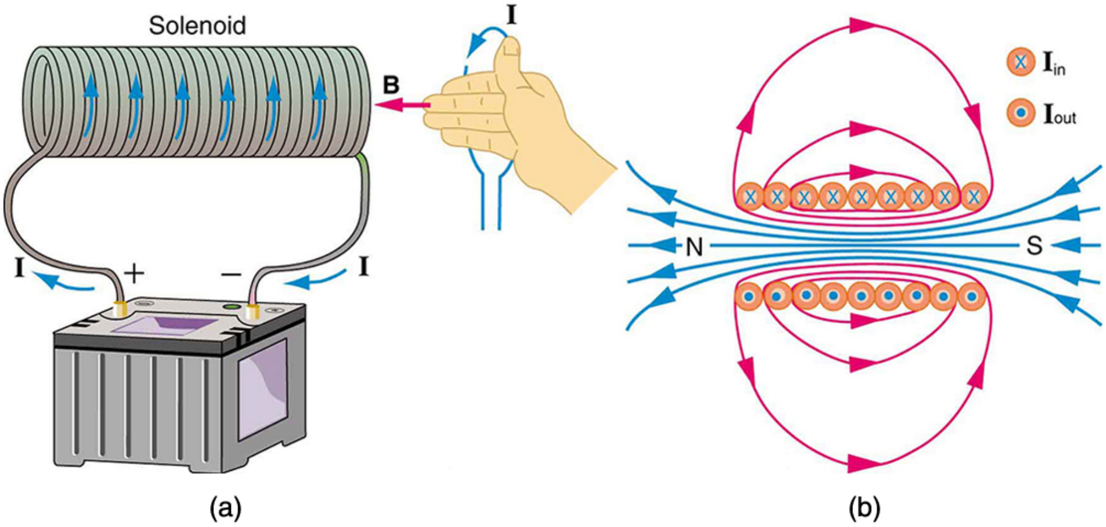


(a) RHR-2 gives the direction of the magnetic field inside and outside a current-carrying loop. (b) More detailed mapping with compasses or with a

Hall probe completes the picture. The field is similar to that of a bar magnet.

Magnetic Field Produced by a Current-Carrying Solenoid

A **solenoid** is a long coil of wire (with many turns or loops, as opposed to a flat loop). Because of its shape, the field inside a solenoid can be very uniform, and also very strong. The field just outside the coils is nearly zero. [link] shows how the field looks and how its direction is given by RHR-2.



(a) Because of its shape, the field inside a solenoid of length l is remarkably uniform in magnitude and direction, as indicated by the straight and uniformly spaced field lines. The field outside the coils is nearly zero. (b) This cutaway shows the magnetic field generated by the current in the solenoid.

The magnetic field inside of a current-carrying solenoid is very uniform in direction and magnitude. Only near the ends does it begin to weaken and change direction. The field outside has similar complexities to flat loops and bar magnets, but the **magnetic field strength inside a solenoid** is simply

Equation:

$$B=\mu_0\text{nI}\ \ (\text{inside a solenoid}),$$

where n is the number of loops per unit length of the solenoid $(n=N/l,$ with N being the number of loops and l the length). Note that B is the field strength anywhere in the uniform region of the interior and not just at the center. Large uniform fields spread over a large volume are possible with solenoids, as [link] implies.

Example:

Calculating Field Strength inside a Solenoid

What is the field inside a 2.00-m-long solenoid that has 2000 loops and carries a 1600-A current?

Strategy

To find the field strength inside a solenoid, we use $B=\mu_0\text{nI}$. First, we note the number of loops per unit length is

Equation:

$$n=\frac{N}{l}=\frac{2000}{2.00\text{ m}}=1000\text{ m}^{-1}=10\text{ cm}^{-1}.$$

Solution

Substituting known values gives

Equation:

$$\begin{aligned}B&=\mu_0\text{nI}=\left(4\pi\times10^{-7}\text{ T}\cdot\text{m/A}\right)\left(1000\text{ m}^{-1}\right)(1600\text{ A})\\&=2.01\text{ T}.\end{aligned}$$

Discussion

This is a large field strength that could be established over a large-diameter solenoid, such as in medical uses of magnetic resonance imaging (MRI). The very large current is an indication that the fields of this strength are not easily achieved, however. Such a large current through 1000 loops squeezed into a meter’s length would produce significant heating. Higher currents can be achieved by using superconducting wires, although this is expensive. There is an upper limit to the current, since the superconducting state is disrupted by very large magnetic fields.

There are interesting variations of the flat coil and solenoid. For example, the toroidal coil used to confine the reactive particles in tokamaks is much like a solenoid bent into a circle. The field inside a toroid is very strong but circular. Charged particles travel in circles, following the field lines, and collide with one another, perhaps inducing fusion. But the charged particles do not cross field lines and escape the toroid. A whole range of coil shapes are used to produce all sorts of magnetic field shapes. Adding ferromagnetic materials produces greater field strengths and can have a significant effect on the shape of the field. Ferromagnetic materials tend to trap magnetic fields (the field lines bend into the ferromagnetic material, leaving weaker fields outside it) and are used as shields for devices that are adversely affected by magnetic fields, including the Earth’s magnetic field.

Note:
PhET Explorations: Generator
Generate electricity with a bar magnet! Discover the physics behind the phenomena by exploring magnets and how you can use them to make a bulb light.

Generator

Section Summary

- The strength of the magnetic field created by current in a long straight wire is given by
  Equation:

  $$B = \frac{\mu_0 I}{2\pi r}(\text{long straight wire}),$$

  where I is the current, r is the shortest distance to the wire, and the constant $\mu_0 = 4\pi \times 10^{-7}\ \text{T}\cdot\text{m/A}$ is the permeability of free space.
- The direction of the magnetic field created by a long straight wire is given by right hand rule 2 (RHR-2): *Point the thumb of the right hand in the direction of current, and the fingers curl in the direction of the magnetic field loops* created by it.
- The magnetic field created by current following any path is the sum (or integral) of the fields due to segments along the path (magnitude and direction as for a straight wire), resulting in a general relationship between current and field known as Ampere’s law.
- The magnetic field strength at the center of a circular loop is given by
  Equation:

  $$B = \frac{\mu_0 I}{2R}(\text{at center of loop}),$$

  where R is the radius of the loop. This equation becomes $B = \mu_0 \text{nI}/(2R)$ for a flat coil of N loops. RHR-2 gives the direction of the field about the loop. A long coil is called a solenoid.
- The magnetic field strength inside a solenoid is
  Equation:

  $$B = \mu_0 \text{nI}\ \ (\text{inside a solenoid}),$$

  where n is the number of loops per unit length of the solenoid. The field inside is very uniform in magnitude and direction.

Conceptual Questions

Exercise:

Problem:

Make a drawing and use RHR-2 to find the direction of the magnetic field of a current loop in a motor (such as in [link]). Then show that the direction of the torque on the loop is the same as produced by like poles repelling and unlike poles attracting.

Glossary

right hand rule 2 (RHR-2)
: a rule to determine the direction of the magnetic field induced by a current-carrying wire: Point the thumb of the right hand in the direction of current, and the fingers curl in the direction of the magnetic field loops

magnetic field strength (magnitude) produced by a long straight current-carrying wire
: defined as $B=\frac{\mu_0 I}{2\pi r}$, where I is the current, r is the shortest distance to the wire, and μ_0 is the permeability of free space

permeability of free space
: the measure of the ability of a material, in this case free space, to support a magnetic field; the constant $\mu_0 = 4\pi \times 10^{-7}\ \mathrm{T \cdot m/A}$

magnetic field strength at the center of a circular loop
: defined as $B=\frac{\mu_0 I}{2R}$ where R is the radius of the loop

solenoid
: a thin wire wound into a coil that produces a magnetic field when an electric current is passed through it

magnetic field strength inside a solenoid
: defined as $B=\mu_0 nI$ where n is the number of loops per unit length of the solenoid $(n = N/l$, with N being the number of loops and l the length)

Biot-Savart law

a physical law that describes the magnetic field generated by an electric current in terms of a specific equation

Ampere’s law

the physical law that states that the magnetic field around an electric current is proportional to the current; each segment of current produces a magnetic field like that of a long straight wire, and the total field of any shape current is the vector sum of the fields due to each segment

Maxwell’s equations

a set of four equations that describe electromagnetic phenomena

Magnetic Force between Two Parallel Conductors

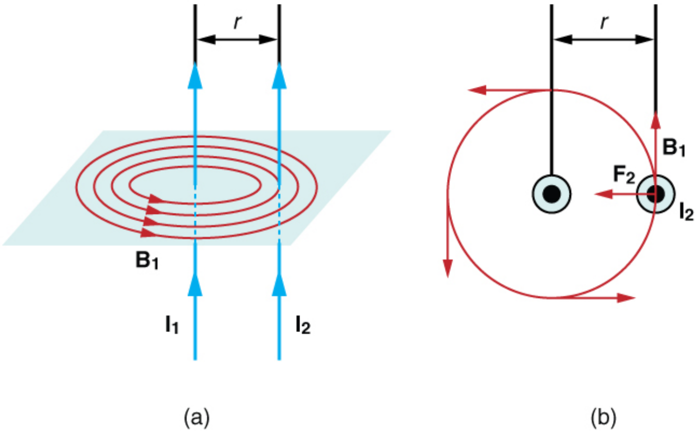
- Describe the effects of the magnetic force between two conductors.
- Calculate the force between two parallel conductors.

You might expect that there are significant forces between current-carrying wires, since ordinary currents produce significant magnetic fields and these fields exert significant forces on ordinary currents. But you might not expect that the force between wires is used to *define* the ampere. It might also surprise you to learn that this force has something to do with why large circuit breakers burn up when they attempt to interrupt large currents.

The force between two long straight and parallel conductors separated by a distance r can be found by applying what we have developed in preceding sections. [link] shows the wires, their currents, the fields they create, and the subsequent forces they exert on one another. Let us consider the field produced by wire 1 and the force it exerts on wire 2 (call the force F_2). The field due to I_1 at a distance r is given to be

Equation:

$$B_1 = \frac{\mu_0 I_1}{2\pi r}.$$



(a) The magnetic field produced by a long straight conductor is perpendicular to a parallel conductor, as indicated by RHR-2. (b) A view from above of the two wires shown in (a), with one magnetic field line shown for each wire. RHR-1 shows that the force between the parallel conductors is attractive when the currents are in the same direction. A similar analysis shows that the force is repulsive between currents in opposite directions.

This field is uniform along wire 2 and perpendicular to it, and so the force F_2 it exerts on wire 2 is given by $F = \text{IlB} \sin \theta$ with $\sin \theta = 1$:

Equation:

$$F_2 = I_2 l B_1.$$

By Newton’s third law, the forces on the wires are equal in magnitude, and so we just write F for the magnitude of F_2. (Note that $F_1 = -F_2$.) Since the wires are very long, it is convenient to think in terms of F/l, the force per unit length. Substituting the expression for B_1 into the last equation and rearranging terms gives

Equation:

$$\frac{F}{l} = \frac{\mu_0 I_1 I_2}{2\pi r}.$$

F/l is the force per unit length between two parallel currents I_1 and I_2 separated by a distance r. The force is attractive if the currents are in the same direction and repulsive if they are in opposite directions.

This force is responsible for the *pinch effect* in electric arcs and plasmas. The force exists whether the currents are in wires or not. In an electric arc, where currents are moving parallel to one another, there is an attraction that squeezes currents into a smaller tube. In large circuit breakers, like those used in neighborhood power distribution systems, the pinch effect can concentrate an arc between plates of a switch trying to break a large current, burn holes, and even ignite the equipment. Another example of the pinch effect is found in the solar plasma, where jets of ionized material, such as solar flares, are shaped by magnetic forces.

The *operational definition of the ampere* is based on the force between current-carrying wires. Note that for parallel wires separated by 1 meter with each carrying 1 ampere, the force per meter is

Equation:

$$\frac{F}{l}=\frac{\left(4\pi\times 10^{-7}\ \mathrm{T}\cdot\mathrm{m/A}\right)(1\ \mathrm{A})^2}{(2\pi)(1\ \mathrm{m})}=2\times 10^{-7}\ \mathrm{N/m}.$$

Since μ_0 is exactly $4\pi\times 10^{-7}\ \mathrm{T}\cdot\mathrm{m/A}$ by definition, and because $1\ \mathrm{T}=1\ \mathrm{N/(A\cdot m)}$, the force per meter is exactly $2\times 10^{-7}\ \mathrm{N/m}$. This is the basis of the operational definition of the ampere.

Note:

The Ampere

The official definition of the ampere is:

One ampere of current through each of two parallel conductors of infinite length, separated by one meter in empty space free of other magnetic fields, causes a force of exactly $2\times 10^{-7}\ \mathrm{N/m}$ on each conductor.

Infinite-length straight wires are impractical and so, in practice, a current balance is constructed with coils of wire separated by a few centimeters. Force is measured to determine current. This also provides us with a method for measuring the coulomb. We measure the charge that flows for a current of one ampere in one second. That is, $1\ \mathrm{C}=1\ \mathrm{A}\cdot\mathrm{s}$. For both the ampere and the coulomb, the method of measuring force between conductors is the most accurate in practice.

Section Summary

- The force between two parallel currents I_1 and I_2, separated by a distance r, has a magnitude per unit length given by
  Equation:

  $$\frac{F}{l} = \frac{\mu_0 I_1 I_2}{2\pi r}.$$

- The force is attractive if the currents are in the same direction, repulsive if they are in opposite directions.

Conceptual Questions

Exercise:

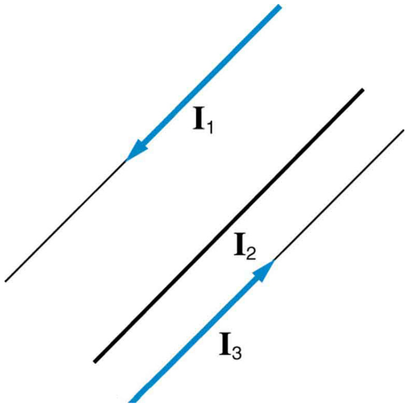
Problem:

Is the force attractive or repulsive between the hot and neutral lines hung from power poles? Why?

Exercise:

Problem:

If you have three parallel wires in the same plane, as in [link], with currents in the outer two running in opposite directions, is it possible for the middle wire to be repelled by both? Attracted by both? Explain.



Three parallel

coplanar wires with currents in the outer two in opposite directions.

Exercise:

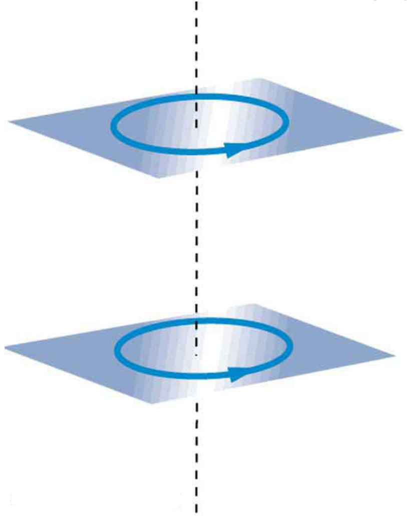
Problem:

Suppose two long straight wires run perpendicular to one another without touching. Does one exert a net force on the other? If so, what is its direction? Does one exert a net torque on the other? If so, what is its direction? Justify your responses by using the right hand rules.

Exercise:

Problem:

Use the right hand rules to show that the force between the two loops in [link] is attractive if the currents are in the same direction and repulsive if they are in opposite directions. Is this consistent with like poles of the loops repelling and unlike poles of the loops attracting? Draw sketches to justify your answers.



Two loops of wire carrying currents

can exert forces
and torques on one
another.

Exercise:

Problem:

If one of the loops in [link] is tilted slightly relative to the other and their currents are in the same direction, what are the directions of the torques they exert on each other? Does this imply that the poles of the bar magnet-like fields they create will line up with each other if the loops are allowed to rotate?

Exercise:

Problem:

Electric field lines can be shielded by the Faraday cage effect. Can we have magnetic shielding? Can we have gravitational shielding?

Problems & Exercises

Exercise:

Problem:

(a) The hot and neutral wires supplying DC power to a light-rail commuter train carry 800 A and are separated by 75.0 cm. What is the magnitude and direction of the force between 50.0 m of these wires? (b) Discuss the practical consequences of this force, if any.

Solution:

(a) 8.53 N, repulsive

(b) This force is repulsive and therefore there is never a risk that the two wires will touch and short circuit.

Exercise:

Problem:

The force per meter between the two wires of a jumper cable being used to start a stalled car is 0.225 N/m. (a) What is the current in the wires, given they are separated by 2.00 cm? (b) Is the force attractive or repulsive?

Exercise:

Problem:

A 2.50-m segment of wire supplying current to the motor of a submerged submarine carries 1000 A and feels a 4.00-N repulsive force from a parallel wire 5.00 cm away. What is the direction and magnitude of the current in the other wire?

Solution:

400 A in the opposite direction

Exercise:

Problem:

The wire carrying 400 A to the motor of a commuter train feels an attractive force of $4.00 \times 10^{-3}\ \mathrm{N/m}$ due to a parallel wire carrying 5.00 A to a headlight. (a) How far apart are the wires? (b) Are the currents in the same direction?

Exercise:

Problem:

An AC appliance cord has its hot and neutral wires separated by 3.00 mm and carries a 5.00-A current. (a) What is the average force per meter between the wires in the cord? (b) What is the maximum force per meter between the wires? (c) Are the forces attractive or repulsive? (d) Do appliance cords need any special design features to compensate for these forces?

Solution:

(a) $1.67 \times 10^{-3}\ \text{N/m}$

(b) $3.33 \times 10^{-3}\ \text{N/m}$

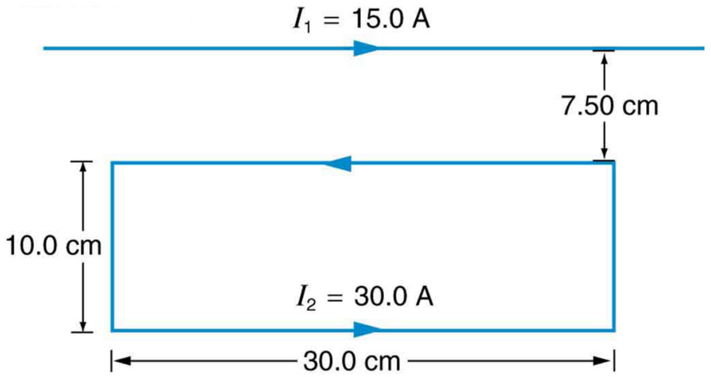
(c) Repulsive

(d) No, these are very small forces

Exercise:

Problem:

[link] shows a long straight wire near a rectangular current loop. What is the direction and magnitude of the total force on the loop?

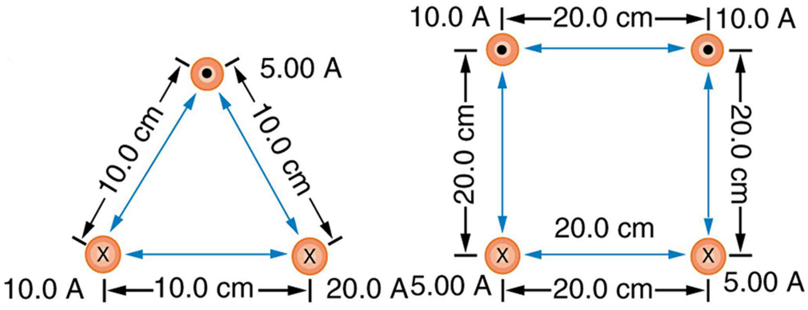


I_1 = 15.0 A
7.50 cm
10.0 cm
I_2 = 30.0 A
30.0 cm

Exercise:

Problem:

Find the direction and magnitude of the force that each wire experiences in [link](a) by, using vector addition.



5.00 A
10.0 cm
10.0 cm
10.0 A
10.0 cm
20.0 A
10.0 A
20.0 cm
10.0 A
20.0 cm
20.0 cm
20.0 cm
5.00 A
20.0 cm
5.00 A

Solution:

(a) Top wire: $2.65 \times 10^{-4}\ \text{N/m s}$, 10.9^{o} to left of up

(b) Lower left wire: $3.61 \times 10^{-4}\ \text{N/m}$, 13.9^{o} down from right

(c) Lower right wire: $3.46 \times 10^{-4}\ \text{N/m}$, 30.0^{o} down from left

Exercise:

Problem:

Find the direction and magnitude of the force that each wire experiences in [link](b), using vector addition.

More Applications of Magnetism

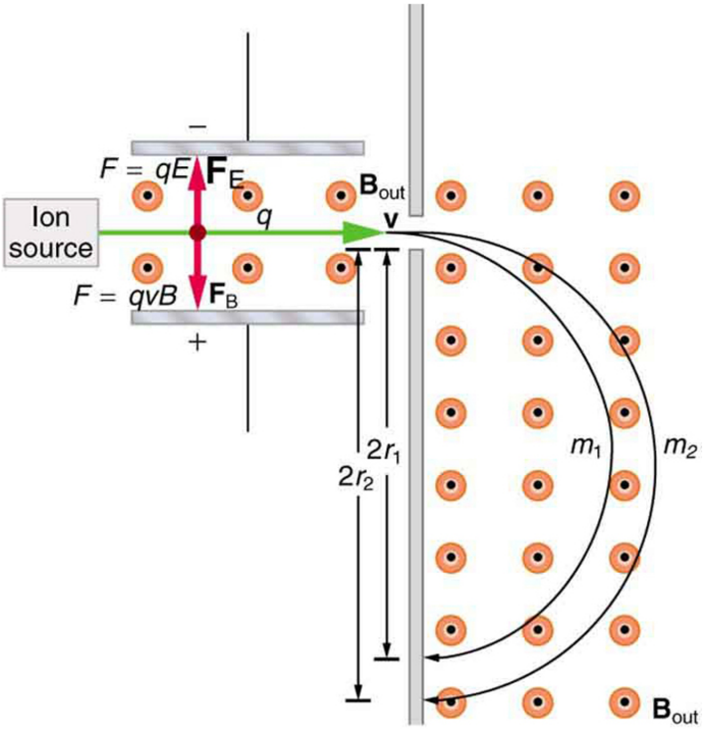
- Describe some applications of magnetism.

Mass Spectrometry

The curved paths followed by charged particles in magnetic fields can be put to use. A charged particle moving perpendicular to a magnetic field travels in a circular path having a radius r.
Equation:

$$r = \frac{\text{mv}}{\text{qB}}$$

It was noted that this relationship could be used to measure the mass of charged particles such as ions. A mass spectrometer is a device that measures such masses. Most mass spectrometers use magnetic fields for this purpose, although some of them have extremely sophisticated designs. Since there are five variables in the relationship, there are many possibilities. However, if v, q, and B can be fixed, then the radius of the path r is simply proportional to the mass m of the charged particle. Let us examine one such mass spectrometer that has a relatively simple design. (See [link].) The process begins with an ion source, a device like an electron gun. The ion source gives ions their charge, accelerates them to some velocity v, and directs a beam of them into the next stage of the spectrometer. This next region is a *velocity selector* that only allows particles with a particular value of v to get through.



−
$F = qE$
$\mathbf{F}_E$
Ion source
q
$\mathbf{B}_{out}$
$\mathbf{v}$
$F = qvB$
$\mathbf{F}_B$
+
$2r_1$
$2r_2$
m_1
m_2
$\mathbf{B}_{out}$

This mass spectrometer uses a velocity selector to fix v so that the radius of the path is proportional to mass.

The velocity selector has both an electric field and a magnetic field, perpendicular to one another, producing forces in opposite directions on the ions. Only those ions for which the forces balance travel in a straight line into the next region. If the forces balance, then the electric force $F = \text{qE}$ equals the magnetic force $F = \text{qvB}$, so that $\text{qE} = \text{qvB}$. Noting that q cancels, we see that

Equation:

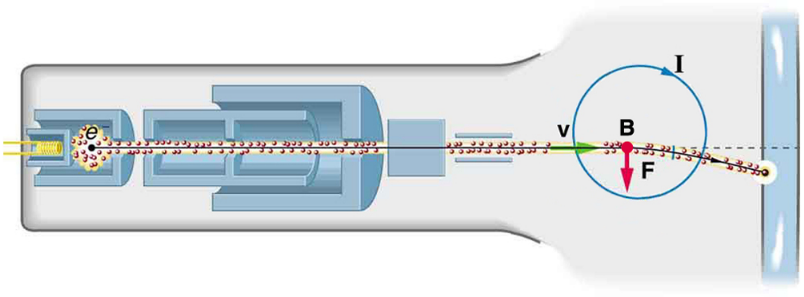
$$v = \frac{E}{B}$$

is the velocity particles must have to make it through the velocity selector, and further, that v can be selected by varying E and B. In the final region, there is only a uniform magnetic field, and so the charged particles move in circular arcs with radii proportional to particle mass. The paths also depend on charge q, but since q is in multiples of electron charges, it is easy to determine and to discriminate between ions in different charge states.

Mass spectrometry today is used extensively in chemistry and biology laboratories to identify chemical and biological substances according to their mass-to-charge ratios. In medicine, mass spectrometers are used to measure the concentration of isotopes used as tracers. Usually, biological molecules such as proteins are very large, so they are broken down into smaller fragments before analyzing. Recently, large virus particles have been analyzed as a whole on mass spectrometers. Sometimes a gas chromatograph or high-performance liquid chromatograph provides an initial separation of the large molecules, which are then input into the mass spectrometer.

Cathode Ray Tubes—CRTs—and the Like

What do non-flat-screen TVs, old computer monitors, x-ray machines, and the 2-mile-long Stanford Linear Accelerator have in common? All of them accelerate electrons, making them different versions of the electron gun. Many of these devices use magnetic fields to steer the accelerated electrons. [link] shows the construction of the type of cathode ray tube (CRT) found in some TVs, oscilloscopes, and old computer monitors. Two pairs of coils are used to steer the electrons, one vertically and the other horizontally, to their desired destination.



The cathode ray tube (CRT) is so named because rays of electrons originate at the cathode in the electron gun. Magnetic coils are used to steer the beam in many CRTs. In this case, the beam is moved down. Another pair of horizontal coils would steer the beam horizontally.

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is one of the most useful and rapidly growing medical imaging tools. It non-invasively produces two-dimensional and three-dimensional images of the body that provide important medical information with none of the hazards of x-rays. MRI is based on an effect called **nuclear magnetic resonance (NMR)** in which an externally applied magnetic field interacts with the nuclei of certain atoms, particularly those of hydrogen (protons). These nuclei possess their own small magnetic fields, similar to those of electrons and the current loops discussed earlier in this chapter.

When placed in an external magnetic field, such nuclei experience a torque that pushes or aligns the nuclei into one of two new energy states—depending on the orientation of its spin (analogous to the N pole and S pole in a bar magnet). Transitions from the lower to higher energy state can be achieved by using an external radio frequency signal to “flip” the orientation of the small magnets. (This is actually a quantum mechanical process. The direction of the nuclear magnetic field is quantized as is energy in the radio waves. We will return to these topics in later chapters.) The specific frequency of the radio waves that are absorbed and reemitted depends sensitively on the type of nucleus, the chemical environment, and the external magnetic field strength. Therefore, this is a *resonance* phenomenon in which *nuclei* in a *magnetic* field act like resonators (analogous to those discussed in the treatment of sound in Oscillatory

Motion and Waves) that absorb and reemit only certain frequencies. Hence, the phenomenon is named *nuclear magnetic resonance (NMR).*

NMR has been used for more than 50 years as an analytical tool. It was formulated in 1946 by F. Bloch and E. Purcell, with the 1952 Nobel Prize in Physics going to them for their work. Over the past two decades, NMR has been developed to produce detailed images in a process now called magnetic resonance imaging (MRI), a name coined to avoid the use of the word “nuclear” and the concomitant implication that nuclear radiation is involved. (It is not.) The 2003 Nobel Prize in Medicine went to P. Lauterbur and P. Mansfield for their work with MRI applications.

The largest part of the MRI unit is a superconducting magnet that creates a magnetic field, typically between 1 and 2 T in strength, over a relatively large volume. MRI images can be both highly detailed and informative about structures and organ functions. It is helpful that normal and non-normal tissues respond differently for slight changes in the magnetic field. In most medical images, the protons that are hydrogen nuclei are imaged. (About 2/3 of the atoms in the body are hydrogen.) Their location and density give a variety of medically useful information, such as organ function, the condition of tissue (as in the brain), and the shape of structures, such as vertebral disks and knee-joint surfaces. MRI can also be used to follow the movement of certain ions across membranes, yielding information on active transport, osmosis, dialysis, and other phenomena. With excellent spatial resolution, MRI can provide information about tumors, strokes, shoulder injuries, infections, etc.

An image requires position information as well as the density of a nuclear type (usually protons). By varying the magnetic field slightly over the volume to be imaged, the resonant frequency of the protons is made to vary with position. Broadcast radio frequencies are swept over an appropriate range and nuclei absorb and reemit them only if the nuclei are in a magnetic field with the correct strength. The imaging receiver gathers information through the body almost point by point, building up a tissue map. The reception of reemitted radio waves as a function of frequency thus gives position information. These “slices” or cross sections through the body are only several mm thick. The intensity of the reemitted radio waves is proportional to the concentration of the nuclear type being flipped, as well as information on the chemical environment in that area of the body. Various techniques are available for enhancing contrast in images and for obtaining more information. Scans called T1, T2, or proton density scans rely on different relaxation mechanisms of nuclei. Relaxation refers to the time it takes for the protons to return to equilibrium after the external field is turned off. This time depends upon tissue type and status (such as inflammation).

While MRI images are superior to x rays for certain types of tissue and have none of the hazards of x rays, they do not completely supplant x-ray images. MRI is less effective than x rays for detecting breaks in bone, for example, and in imaging breast tissue, so the two diagnostic tools complement each other. MRI images are also expensive compared to simple x-ray images and tend to be used most often where they supply information not readily obtained from x rays. Another disadvantage of MRI is that the patient is totally enclosed with detectors close to the body for about 30 minutes or more, leading to claustrophobia. It is also difficult for the obese patient to be in the magnet tunnel. New “open-MRI” machines are now available in which the magnet does not completely surround the patient.

Over the last decade, the development of much faster scans, called “functional MRI” (fMRI), has allowed us to map the functioning of various regions in the brain responsible for thought and motor control. This technique measures the change in blood flow for activities (thought, experiences, action) in the brain. The nerve cells increase their consumption of oxygen when active. Blood hemoglobin releases oxygen to active nerve cells and has somewhat different magnetic properties when oxygenated than when deoxygenated. With MRI, we can measure this and detect a blood oxygen-dependent signal. Most of the brain scans today use fMRI.

Other Medical Uses of Magnetic Fields

Currents in nerve cells and the heart create magnetic fields like any other currents. These can be measured but with some difficulty since their strengths are about 10^{-6} to 10^{-8} *less* than the Earth’s magnetic field. Recording of the heart’s magnetic field as it beats is called a

magnetocardiogram (MCG), while measurements of the brain’s magnetic field is called a **magnetoencephalogram (MEG)**. Both give information that differs from that obtained by measuring the electric fields of these organs (ECGs and EEGs), but they are not yet of sufficient importance to make these difficult measurements common.

In both of these techniques, the sensors do not touch the body. MCG can be used in fetal studies, and is probably more sensitive than echocardiography. MCG also looks at the heart’s electrical activity whose voltage output is too small to be recorded by surface electrodes as in EKG. It has the potential of being a rapid scan for early diagnosis of cardiac ischemia (obstruction of blood flow to the heart) or problems with the fetus.

MEG can be used to identify abnormal electrical discharges in the brain that produce weak magnetic signals. Therefore, it looks at brain activity, not just brain structure. It has been used for studies of Alzheimer’s disease and epilepsy. Advances in instrumentation to measure very small magnetic fields have allowed these two techniques to be used more in recent years. What is used is a sensor called a SQUID, for superconducting quantum interference device. This operates at liquid helium temperatures and can measure magnetic fields thousands of times smaller than the Earth’s.

Finally, there is a burgeoning market for magnetic cures in which magnets are applied in a variety of ways to the body, from magnetic bracelets to magnetic mattresses. The best that can be said for such practices is that they are apparently harmless, unless the magnets get close to the patient’s computer or magnetic storage disks. Claims are made for a broad spectrum of benefits from cleansing the blood to giving the patient more energy, but clinical studies have not verified these claims, nor is there an identifiable mechanism by which such benefits might occur.

Note:
PhET Explorations: Magnet and Compass
Ever wonder how a compass worked to point you to the Arctic? Explore the interactions between a compass and bar magnet, and then add the Earth and find the surprising answer! Vary the magnet's strength, and see how

things change both inside and outside. Use the field meter to measure how the magnetic field changes. https://archive.cnx.org/specials/5ca3e2cc-ae74-11e5-b6d3-f3c228f04b5c/magnet-and-compass/#sim-bar-magnet

Section Summary

- Crossed (perpendicular) electric and magnetic fields act as a velocity filter, giving equal and opposite forces on any charge with velocity perpendicular to the fields and of magnitude
  Equation:
  $$v=\frac{E}{B}.$$

Conceptual Questions

Exercise:

Problem:

Measurements of the weak and fluctuating magnetic fields associated with brain activity are called magnetoencephalograms (MEGs). Do the brain’s magnetic fields imply coordinated or uncoordinated nerve impulses? Explain.

Exercise:

Problem:

Discuss the possibility that a Hall voltage would be generated on the moving heart of a patient during MRI imaging. Also discuss the same effect on the wires of a pacemaker. (The fact that patients with pacemakers are not given MRIs is significant.)

Exercise:

Problem:

A patient in an MRI unit turns his head quickly to one side and experiences momentary dizziness and a strange taste in his mouth. Discuss the possible causes.

Exercise:

Problem:

You are told that in a certain region there is either a uniform electric or magnetic field. What measurement or observation could you make to determine the type? (Ignore the Earth’s magnetic field.)

Exercise:

Problem:

An example of magnetohydrodynamics (MHD) comes from the flow of a river (salty water). This fluid interacts with the Earth’s magnetic field to produce a potential difference between the two river banks. How would you go about calculating the potential difference?

Exercise:

Problem:

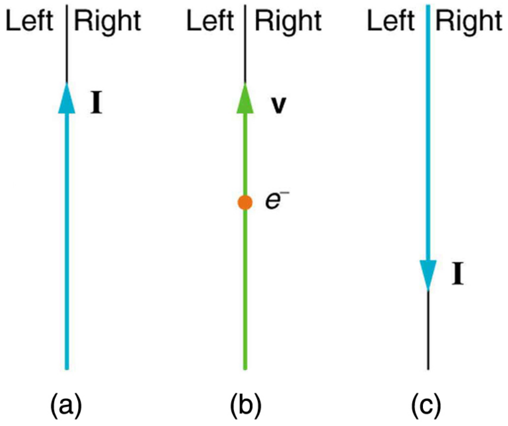
Draw gravitational field lines between 2 masses, electric field lines between a positive and a negative charge, electric field lines between 2 positive charges and magnetic field lines around a magnet. Qualitatively describe the differences between the fields and the entities responsible for the field lines.

Problems & Exercises

Exercise:

Problem:

Indicate whether the magnetic field created in each of the three situations shown in [link] is into or out of the page on the left and right of the current.


Left
Right
I
Left
Right
v
e^-
Left
Right
I
(a)
(b)
(c)

Solution:

(a) right-into page, left-out of page

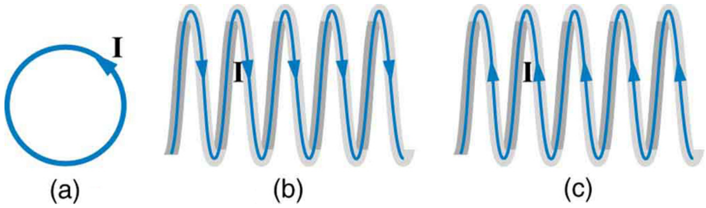
(b) right-out of page, left-into page

(c) right-out of page, left-into page

Exercise:

Problem:

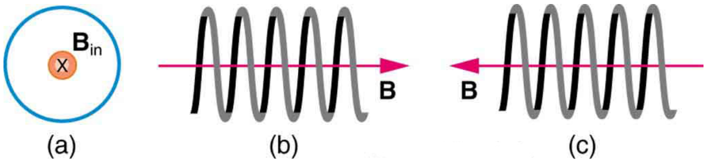
What are the directions of the fields in the center of the loop and coils shown in [link]?



Exercise:

Problem:

What are the directions of the currents in the loop and coils shown in [link]?



(a) (b) (c)

Solution:

(a) clockwise

(b) clockwise as seen from the left

(c) clockwise as seen from the right

Exercise:

Problem:

To see why an MRI utilizes iron to increase the magnetic field created by a coil, calculate the current needed in a 400-loop-per-meter circular coil 0.660 m in radius to create a 1.20-T field (typical of an MRI instrument) at its center with no iron present. The magnetic field of a proton is approximately like that of a circular current loop 0.650×10^{-15} m in radius carrying 1.05×10^{4} A. What is the field at the center of such a loop?

Solution:

1.01×10^{13} T

Exercise:

Problem:

Inside a motor, 30.0 A passes through a 250-turn circular loop that is 10.0 cm in radius. What is the magnetic field strength created at its center?

Exercise:

Problem:

Nonnuclear submarines use batteries for power when submerged. (a) Find the magnetic field 50.0 cm from a straight wire carrying 1200 A from the batteries to the drive mechanism of a submarine. (b) What is the field if the wires to and from the drive mechanism are side by side? (c) Discuss the effects this could have for a compass on the submarine that is not shielded.

Solution:

(a) $4.80 \times 10^{-4}\ \text{T}$

(b) Zero

(c) If the wires are not paired, the field is about 10 times stronger than Earth’s magnetic field and so could severely disrupt the use of a compass.

Exercise:

Problem:

How strong is the magnetic field inside a solenoid with 10,000 turns per meter that carries 20.0 A?

Exercise:

Problem:

What current is needed in the solenoid described in [link] to produce a magnetic field 10^4 times the Earth’s magnetic field of $5.00 \times 10^{-5}\ \text{T}$?

Solution:

39.8 A

Exercise:

Problem:

How far from the starter cable of a car, carrying 150 A, must you be to experience a field less than the Earth’s $(5.00 \times 10^{-5} \text{ T})$? Assume a long straight wire carries the current. (In practice, the body of your car shields the dashboard compass.)

Exercise:

Problem:

Measurements affect the system being measured, such as the current loop in [link]. (a) Estimate the field the loop creates by calculating the field at the center of a circular loop 20.0 cm in diameter carrying 5.00 A. (b) What is the smallest field strength this loop can be used to measure, if its field must alter the measured field by less than 0.0100%?

Solution:

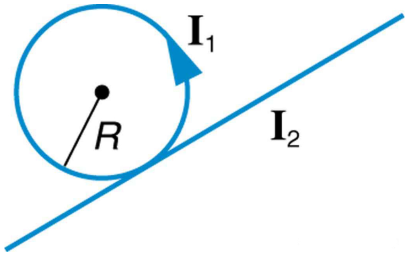
(a) $3.14 \times 10^{-5} \text{ T}$

(b) 0.314 T

Exercise:

Problem:

[link] shows a long straight wire just touching a loop carrying a current I_1. Both lie in the same plane. (a) What direction must the current I_2 in the straight wire have to create a field at the center of the loop in the direction opposite to that created by the loop? (b) What is the ratio of I_1/I_2 that gives zero field strength at the center of the loop? (c) What is the direction of the field directly above the loop under this circumstance?



Exercise:

Problem:

Find the magnitude and direction of the magnetic field at the point equidistant from the wires in [link](a), using the rules of vector addition to sum the contributions from each wire.

Solution:

$7.55 \times 10^{-5}\ \text{T},\ 23.4^{\text{o}}$

Exercise:

Problem:

Find the magnitude and direction of the magnetic field at the point equidistant from the wires in [link](b), using the rules of vector addition to sum the contributions from each wire.

Exercise:

Problem:

What current is needed in the top wire in [link](a) to produce a field of zero at the point equidistant from the wires, if the currents in the bottom two wires are both 10.0 A into the page?

Solution:

10.0 A

Exercise:

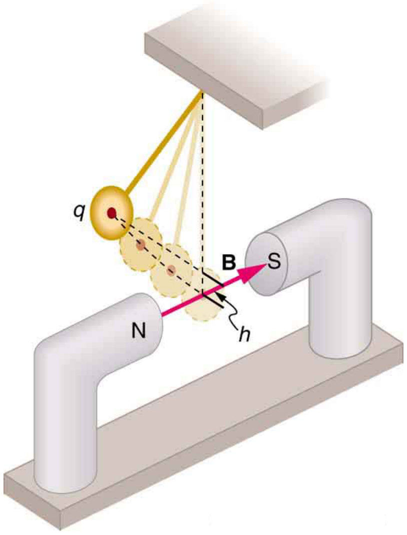
Problem:

Calculate the size of the magnetic field 20 m below a high voltage power line. The line carries 450 MW at a voltage of 300,000 V.

Exercise:

Problem: **Integrated Concepts**

(a) A pendulum is set up so that its bob (a thin copper disk) swings between the poles of a permanent magnet as shown in [link]. What is the magnitude and direction of the magnetic force on the bob at the lowest point in its path, if it has a positive $0.250\ \mu\text{C}$ charge and is released from a height of 30.0 cm above its lowest point? The magnetic field strength is 1.50 T. (b) What is the acceleration of the bob at the bottom of its swing if its mass is 30.0 grams and it is hung from a flexible string? Be certain to include a free-body diagram as part of your analysis.



Solution:

(a) 9.09×10^{-7} N upward

(b) $3.03 \times 10^{-5}\ \text{m/s}^2$

Exercise:

Problem: Integrated Concepts

(a) What voltage will accelerate electrons to a speed of $6.00 \times 10^{-7}\ \text{m/s}$? (b) Find the radius of curvature of the path of a *proton* accelerated through this potential in a 0.500-T field and compare this with the radius of curvature of an electron accelerated through the same potential.

Exercise:

Problem: Integrated Concepts

Find the radius of curvature of the path of a 25.0-MeV proton moving perpendicularly to the 1.20-T field of a cyclotron.

Solution:

60.2 cm

Exercise:

Problem: Integrated Concepts

To construct a nonmechanical water meter, a 0.500-T magnetic field is placed across the supply water pipe to a home and the Hall voltage is recorded. (a) Find the flow rate in liters per second through a 3.00-cm-diameter pipe if the Hall voltage is 60.0 mV. (b) What would the Hall voltage be for the same flow rate through a 10.0-cm-diameter pipe with the same field applied?

Exercise:

Problem: Integrated Concepts

(a) Using the values given for an MHD drive in [link], and assuming the force is uniformly applied to the fluid, calculate the pressure created in N/m^2. (b) Is this a significant fraction of an atmosphere?

Solution:

(a) $1.02 \times 10^3 \text{ N/m}^2$

(b) Not a significant fraction of an atmosphere

Exercise:

Problem: **Integrated Concepts**

(a) Calculate the maximum torque on a 50-turn, 1.50 cm radius circular current loop carrying $50\ \mu\text{A}$ in a 0.500-T field. (b) If this coil is to be used in a galvanometer that reads $50\ \mu\text{A}$ full scale, what force constant spring must be used, if it is attached 1.00 cm from the axis of rotation and is stretched by the 60^o arc moved?

Exercise:

Problem: **Integrated Concepts**

A current balance used to define the ampere is designed so that the current through it is constant, as is the distance between wires. Even so, if the wires change length with temperature, the force between them will change. What percent change in force per degree will occur if the wires are copper?

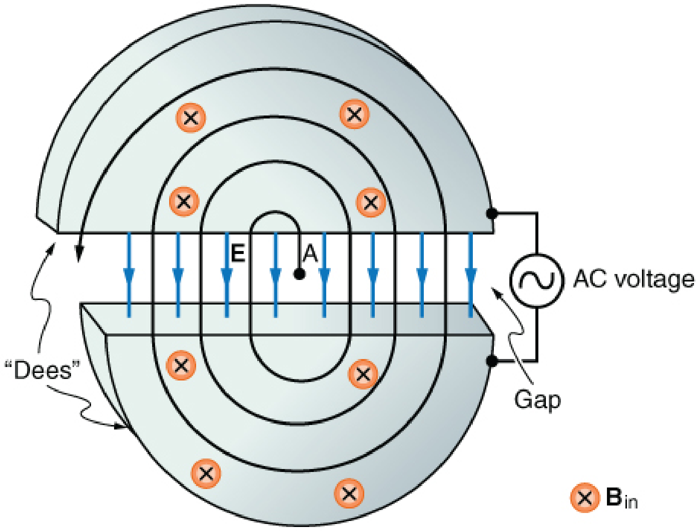
Solution:

$17.0 \times 10^{-4}\%/^\text{o}\text{C}$

Exercise:

Problem:**Integrated Concepts**

(a) Show that the period of the circular orbit of a charged particle moving perpendicularly to a uniform magnetic field is $T = 2\pi m/(\text{qB})$. (b) What is the frequency f? (c) What is the angular velocity ω? Note that these results are independent of the velocity and radius of the orbit and, hence, of the energy of the particle. ([link].)



Cyclotrons accelerate charged particles orbiting in a magnetic field by placing an AC voltage on the metal Dees, between which the particles move, so that energy is added twice each orbit. The frequency is constant, since it is independent of the particle energy—the radius of the orbit simply increases with energy until the particles approach the edge and are extracted for various experiments and applications.

Exercise:

Problem: Integrated Concepts

A cyclotron accelerates charged particles as shown in [link]. Using the results of the previous problem, calculate the frequency of the accelerating voltage needed for a proton in a 1.20-T field.

Solution:

18.3 MHz

Exercise:

Problem: Integrated Concepts

(a) A 0.140-kg baseball, pitched at 40.0 m/s horizontally and perpendicular to the Earth’s horizontal $5.00 \times 10^{-5}\ \text{T}$ field, has a 100-nC charge on it. What distance is it deflected from its path by the magnetic force, after traveling 30.0 m horizontally? (b) Would you suggest this as a secret technique for a pitcher to throw curve balls?

Exercise:

Problem: Integrated Concepts

(a) What is the direction of the force on a wire carrying a current due east in a location where the Earth’s field is due north? Both are parallel to the ground. (b) Calculate the force per meter if the wire carries 20.0 A and the field strength is $3.00 \times 10^{-5}\ \text{T}$. (c) What diameter copper wire would have its weight supported by this force? (d) Calculate the resistance per meter and the voltage per meter needed.

Solution:

(a) Straight up

(b) $6.00 \times 10^{-4}\ \text{N/m}$

(c) 94.1 μm

(d)2.47 Ω/m, 49.4 V/m

Exercise:

Problem: Integrated Concepts

One long straight wire is to be held directly above another by repulsion between their currents. The lower wire carries 100 A and the wire 7.50 cm above it is 10-gauge (2.588 mm diameter) copper wire. (a) What current must flow in the upper wire, neglecting the Earth's field? (b) What is the smallest current if the Earth's $3.00 \times 10^{-5}\ \mathrm{T}$ field is parallel to the ground and is not neglected? (c) Is the supported wire in a stable or unstable equilibrium if displaced vertically? If displaced horizontally?

Exercise:

Problem: Unreasonable Results

(a) Find the charge on a baseball, thrown at 35.0 m/s perpendicular to the Earth's $5.00 \times 10^{-5}\ \mathrm{T}$ field, that experiences a 1.00-N magnetic force. (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

(a) 571 C

(b) Impossible to have such a large separated charge on such a small object.

(c) The 1.00-N force is much too great to be realistic in the Earth's field.

Exercise:

Problem: Unreasonable Results

A charged particle having mass $6.64 \times 10^{-27}\ \mathrm{kg}$ (that of a helium atom) moving at $8.70 \times 10^{5}\ \mathrm{m/s}$ perpendicular to a 1.50-T magnetic field travels in a circular path of radius 16.0 mm. (a) What is the charge of the particle? (b) What is unreasonable about this result? (c) Which assumptions are responsible?

Exercise:

Problem:Unreasonable Results

An inventor wants to generate 120-V power by moving a 1.00-m-long wire perpendicular to the Earth’s $5.00 \times 10^{-5}\ \mathrm{T}$ field. (a) Find the speed with which the wire must move. (b) What is unreasonable about this result? (c) Which assumption is responsible?

Solution:

(a) $2.40 \times 10^{6}\ \mathrm{m/s}$

(b) The speed is too high to be practical $\leq$ 1% speed of light

(c) The assumption that you could reasonably generate such a voltage with a single wire in the Earth’s field is unreasonable

Exercise:

Problem: Unreasonable Results

Frustrated by the small Hall voltage obtained in blood flow measurements, a medical physicist decides to increase the applied magnetic field strength to get a 0.500-V output for blood moving at 30.0 cm/s in a 1.50-cm-diameter vessel. (a) What magnetic field strength is needed? (b) What is unreasonable about this result? (c) Which premise is responsible?

Exercise:

Problem:Unreasonable Results

A surveyor 100 m from a long straight 200-kV DC power line suspects that its magnetic field may equal that of the Earth and affect compass readings. (a) Calculate the current in the wire needed to create a $5.00 \times 10^{-5}\ \mathrm{T}$ field at this distance. (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

(a) 25.0 kA

(b) This current is unreasonably high. It implies a total power delivery in the line of 50.0x10^9 W, which is much too high for standard transmission lines.

(c)100 meters is a long distance to obtain the required field strength. Also coaxial cables are used for transmission lines so that there is virtually no field for DC power lines, because of cancellation from opposing currents. The surveyor’s concerns are not a problem for his magnetic field measurements.

Exercise:

Problem:Construct Your Own Problem

Consider a mass separator that applies a magnetic field perpendicular to the velocity of ions and separates the ions based on the radius of curvature of their paths in the field. Construct a problem in which you calculate the magnetic field strength needed to separate two ions that differ in mass, but not charge, and have the same initial velocity. Among the things to consider are the types of ions, the velocities they can be given before entering the magnetic field, and a reasonable value for the radius of curvature of the paths they follow. In addition, calculate the separation distance between the ions at the point where they are detected.

Exercise:

Problem: Construct Your Own Problem

Consider using the torque on a current-carrying coil in a magnetic field to detect relatively small magnetic fields (less than the field of the Earth, for example). Construct a problem in which you calculate the maximum torque on a current-carrying loop in a magnetic field. Among the things to be considered are the size of the coil, the number of loops it has, the current you pass through the coil, and the size of the field you wish to detect. Discuss whether the torque produced is large enough to be effectively measured. Your instructor may also wish for you to consider the effects, if any, of the field produced by the coil on the surroundings that could affect detection of the small field.

Glossary

magnetic resonance imaging (MRI)	a medical imaging technique that uses magnetic fields create detailed images of internal tissues and organs
nuclear magnetic resonance (NMR)	a phenomenon in which an externally applied magnetic field interacts with the nuclei of certain atoms
magnetocardiogram (MCG)	a recording of the heart’s magnetic field as it beats
magnetoencephalogram (MEG)	a measurement of the brain’s magnetic field

Introduction to Electromagnetic Induction, AC Circuits and Electrical Technologies

class="introduction"

These wind turbines in the Thames Estuary in the UK are an example of induction at work. Wind pushes the blades of the turbine, spinning a shaft attached to magnets. The magnets spin around a conductive coil, inducing an electric current in the coil, and eventually feeding the electrical grid. (credit: modification of work by Petr Kratochvil)

Nature’s displays of symmetry are beautiful and alluring. A butterfly’s wings exhibit an appealing symmetry in a complex system. (See [link].) The laws of physics display symmetries at the most basic level—these symmetries are a source of wonder and imply deeper meaning. Since we place a high value on symmetry, we look for it when we explore nature. The remarkable thing is that we find it.

Physics, like this butterfly, has inherent symmetries. (credit: Thomas Bresson)

The hint of symmetry between electricity and magnetism found in the preceding chapter will be elaborated upon in this chapter. Specifically, we know that a current creates a magnetic field. If nature is symmetric here, then perhaps a magnetic field can create a current. The Hall effect is a voltage caused by a magnetic force. That voltage could drive a current. Historically, it was very shortly after Oersted discovered currents cause magnetic fields that other scientists asked the following question: Can magnetic fields cause currents? The answer was soon found by experiment to be yes. In 1831, some 12 years after Oersted's discovery, the English scientist Michael Faraday (1791–1862) and the American scientist Joseph Henry (1797–1878) independently demonstrated that magnetic fields can produce currents. The basic process of generating emfs (electromotive force) and, hence, currents with magnetic fields is known as **induction**; this process is also called magnetic induction to distinguish it from charging by induction, which utilizes the Coulomb force.

Today, currents induced by magnetic fields are essential to our technological society. The ubiquitous generator—found in automobiles, on bicycles, in nuclear power plants, and so on—uses magnetism to generate current. Other devices that use magnetism to induce currents include pickup coils in electric guitars, transformers of every size, certain microphones, airport security gates, and damping mechanisms on sensitive chemical balances. Not so familiar perhaps, but important nevertheless, is that the behavior of AC circuits depends strongly on the effect of magnetic fields on currents.

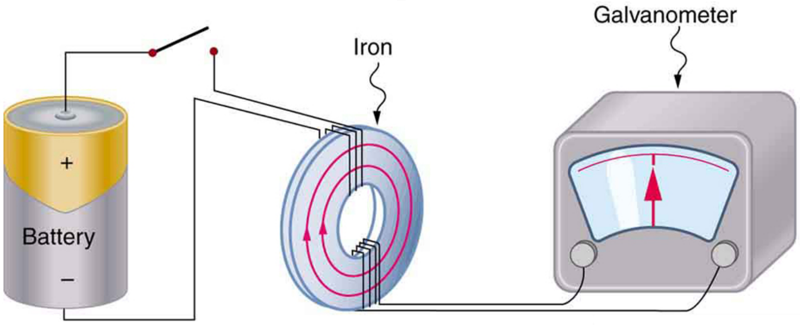
Glossary

induction	(magnetic induction) the creation of emfs and hence currents by magnetic fields

Induced Emf and Magnetic Flux

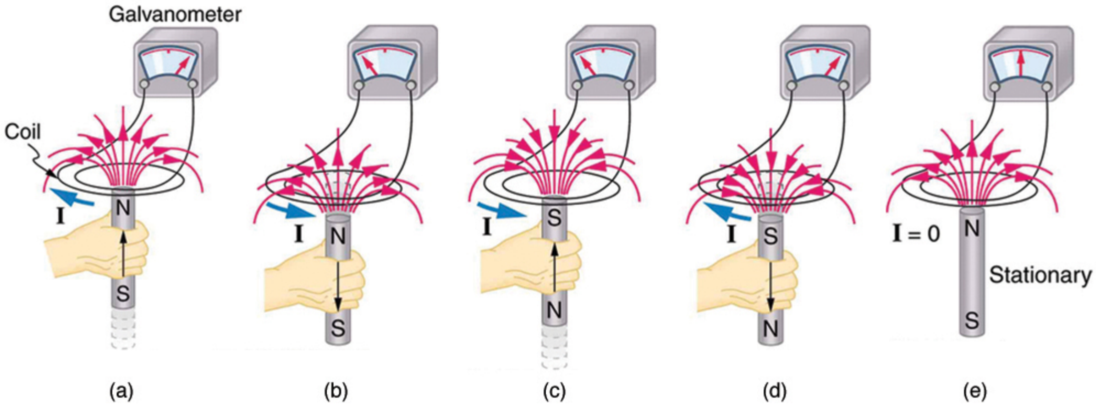
- Calculate the flux of a uniform magnetic field through a loop of arbitrary orientation.
- Describe methods to produce an electromotive force (emf) with a magnetic field or magnet and a loop of wire.

The apparatus used by Faraday to demonstrate that magnetic fields can create currents is illustrated in [link]. When the switch is closed, a magnetic field is produced in the coil on the top part of the iron ring and transmitted to the coil on the bottom part of the ring. The galvanometer is used to detect any current induced in the coil on the bottom. It was found that each time the switch is closed, the galvanometer detects a current in one direction in the coil on the bottom. (You can also observe this in a physics lab.) Each time the switch is opened, the galvanometer detects a current in the opposite direction. Interestingly, if the switch remains closed or open for any length of time, there is no current through the galvanometer. *Closing and opening the switch* induces the current. It is the *change* in magnetic field that creates the current. More basic than the current that flows is the emfthat causes it. The current is a result of an *emf induced by a changing magnetic field*, whether or not there is a path for current to flow.



Faraday’s apparatus for demonstrating that a magnetic field can produce a current. A change in the field produced by the top coil induces an emf and, hence, a current in the bottom coil. When the switch is opened and closed, the galvanometer registers currents in opposite directions. No current flows through the galvanometer when the switch remains closed or open.

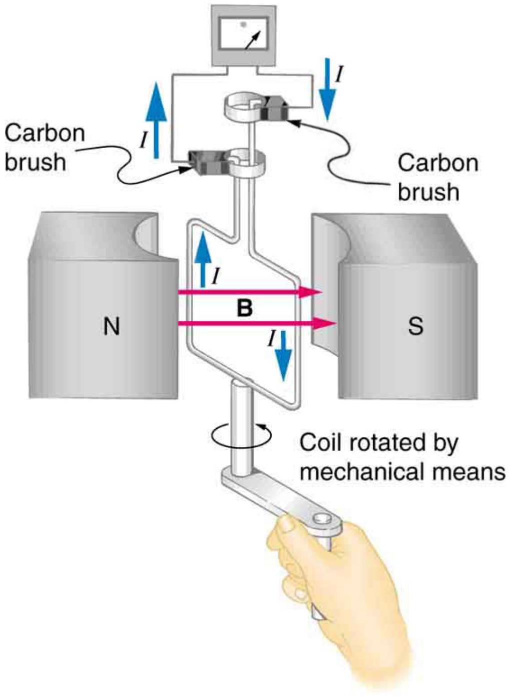
An experiment easily performed and often done in physics labs is illustrated in [link]. An emf is induced in the coil when a bar magnet is pushed in and out of it. Emfs of opposite signs are produced by motion in opposite directions, and the emfs are also reversed by reversing poles. The same results are produced if the coil is moved rather than the magnet—it is the relative motion that is important. The faster the motion, the greater the emf, and there is no emf when the magnet is stationary relative to the coil.



Movement of a magnet relative to a coil produces emfs as shown. The same emfs are produced if the coil is moved relative to the magnet. The greater the speed, the greater the magnitude of the emf, and the emf is zero when there is no motion.

The method of inducing an emf used in most electric generators is shown in [link]. A coil is rotated in a magnetic field, producing an alternating current emf, which depends on rotation rate and other factors that will be explored in later sections. Note that the generator is remarkably similar in construction to a motor (another symmetry).

$$\Phi = \mathrm{BA}\cos\theta,$$


Carbon brush
I
I
Carbon brush
I
B
N
S
I
Coil rotated by mechanical means

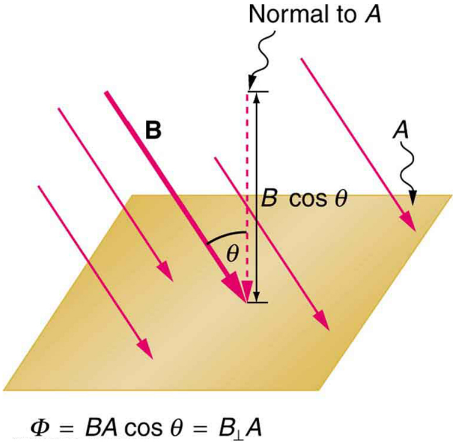
Rotation of a coil in a magnetic field produces an emf. This is the basic construction of a generator, where work done to turn the coil is converted to electric energy. Note the generator is very similar in construction to a motor.

So we see that changing the magnitude or direction of a magnetic field produces an emf. Experiments revealed that there is a crucial quantity called the **magnetic flux**, Φ , given by

Equation:

$$\Phi = \mathrm{BA}\cos\theta,$$

where B is the magnetic field strength over an area A, at an angle θ with the perpendicular to the area as shown in [link]. **Any change in magnetic flux Φ induces an emf.** This process is defined to be **electromagnetic induction**. Units of magnetic flux Φ are $\mathrm{T} \cdot \mathrm{m}^2$. As seen in [link], $B\ \cos\ \theta = B_{\perp}$, which is the component of B perpendicular to the area A. Thus magnetic flux is $\Phi = B_{\perp}A$, the product of the area and the component of the magnetic field perpendicular to it.



Magnetic flux Φ is related to the magnetic field and the area over which it exists. The flux $\Phi = \mathrm{BA}\ \cos\ \theta$ is related to induction; any change in Φ induces an emf.

All induction, including the examples given so far, arises from some change in magnetic flux Φ. For example, Faraday changed B and hence Φ when opening and closing the switch in his apparatus (shown in [link]). This is also true for the bar magnet and coil shown in [link]. When rotating the coil of a generator, the angle θ and, hence, Φ is changed. Just how great an emf and what direction it takes depend on the change in Φ and how rapidly the change is made, as examined in the next section.

Section Summary

- The crucial quantity in induction is magnetic flux Φ, defined to be $\Phi = \text{BA}\cos\theta$, where B is the magnetic field strength over an area A at an angle θ with the perpendicular to the area.
- Units of magnetic flux Φ are $\text{T} \cdot \text{m}^2$.
- Any change in magnetic flux Φ induces an emf—the process is defined to be electromagnetic induction.

Conceptual Questions

Exercise:

Problem:

How do the multiple-loop coils and iron ring in the version of Faraday’s apparatus shown in [link] enhance the observation of induced emf?

Exercise:

Problem:

When a magnet is thrust into a coil as in [link](a), what is the direction of the force exerted by the coil on the magnet? Draw a diagram showing the direction of the current induced in the coil and the magnetic field it produces, to justify your response. How does the magnitude of the force depend on the resistance of the galvanometer?

Exercise:

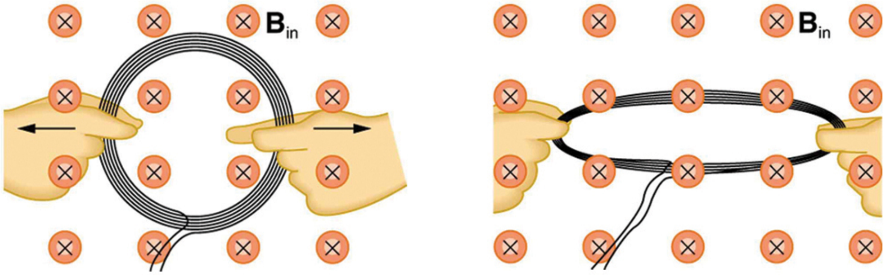
Problem:

Explain how magnetic flux can be zero when the magnetic field is not zero.

Exercise:

Problem:

Is an emf induced in the coil in [link] when it is stretched? If so, state why and give the direction of the induced current.



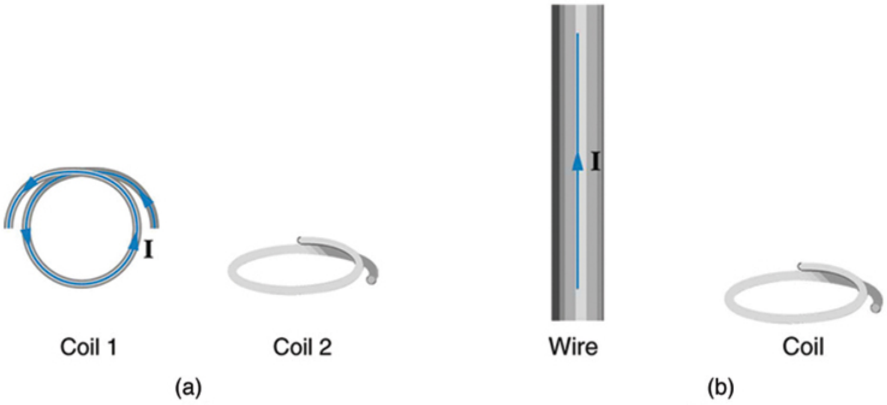
A circular coil of wire is stretched in a magnetic field.

Problems & Exercises

Exercise:

Problem:

What is the value of the magnetic flux at coil 2 in [link] due to coil 1?



(a) The planes of the two coils are perpendicular. (b) The wire is perpendicular to the plane of the coil.

Solution:

Zero

Exercise:

Problem:

What is the value of the magnetic flux through the coil in [link](b) due to the wire?

Glossary

magnetic flux
: the amount of magnetic field going through a particular area, calculated with $\Phi = \mathrm{BA}\cos\theta$ where B is the magnetic field strength over an area A at an angle θ with the perpendicular to the area

electromagnetic induction
: the process of inducing an emf (voltage) with a change in magnetic flux

Faraday’s Law of Induction: Lenz’s Law

- Calculate emf, current, and magnetic fields using Faraday’s Law.
- Explain the physical results of Lenz’s Law

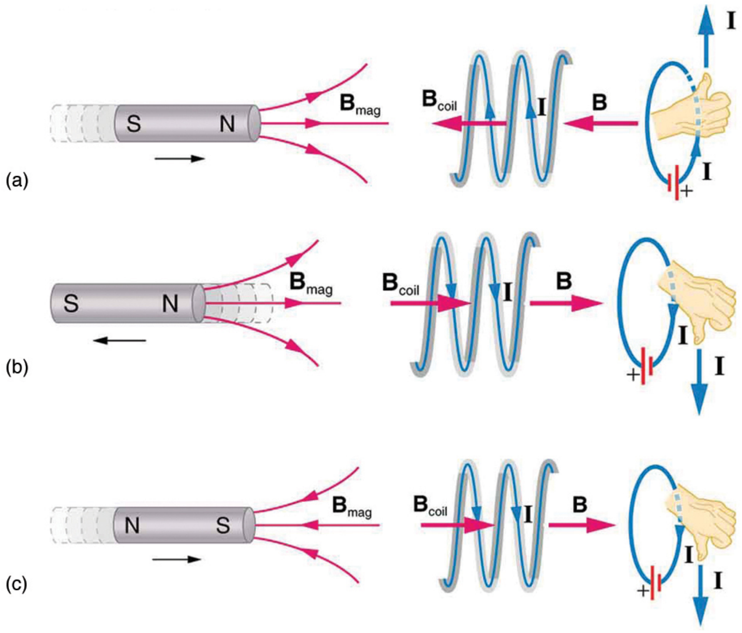
Faraday’s and Lenz’s Law

Faraday’s experiments showed that the emf induced by a change in magnetic flux depends on only a few factors. First, emf is directly proportional to the change in flux $\Delta \mathit{\Phi}$. Second, emf is greatest when the change in time Δt is smallest—that is, emf is inversely proportional to Δt. Finally, if a coil has N turns, an emf will be produced that is N times greater than for a single coil, so that emf is directly proportional to N. The equation for the emf induced by a change in magnetic flux is
Equation:

$$\text{emf} = -N\frac{\Delta \mathit{\Phi}}{\Delta t}.$$

This relationship is known as **Faraday’s law of induction**. The units for emf are volts, as is usual.

The minus sign in Faraday’s law of induction is very important. The minus means that *the emf creates a current I and magnetic field B that oppose the change in flux* $\Delta \mathit{\Phi}$*—this is known as Lenz’s law*. The direction (given by the minus sign) of the emfis so important that it is called **Lenz’s law** after the Russian Heinrich Lenz (1804–1865), who, like Faraday and Henry,independently investigated aspects of induction. Faraday was aware of the direction, but Lenz stated it so clearly that he is credited for its discovery. (See [link].)



(a) When this bar magnet is thrust into the coil, the strength of the magnetic field increases in the coil. The current induced in the coil creates another field, in the opposite direction of the bar magnet's to oppose the increase. This is one aspect of *Lenz's law—induction opposes any change in flux*. (b) and (c) are two other situations. Verify for yourself that the direction of the induced B_{coil} shown indeed opposes the change in flux and that the current direction shown is consistent with RHR-2.

Note:

Problem-Solving Strategy for Lenz's Law

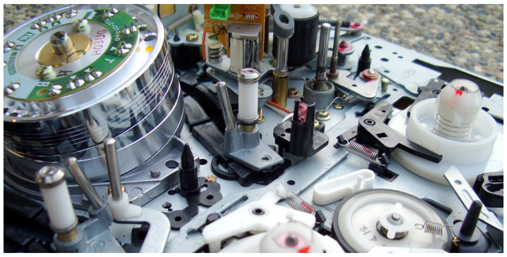
To use Lenz's law to determine the directions of the induced magnetic fields, currents, and emfs:

1. Make a sketch of the situation for use in visualizing and recording directions.
2. Determine the direction of the magnetic field B.
3. Determine whether the flux is increasing or decreasing.
4. Now determine the direction of the induced magnetic field B. It opposes the *change* in flux by adding or subtracting from the original field.
5. Use RHR-2 to determine the direction of the induced current I that is responsible for the induced magnetic field B.
6. The direction (or polarity) of the induced emf will now drive a current in this direction and can be represented as current emerging from the positive terminal of the emf and returning to its negative terminal.

For practice, apply these steps to the situations shown in [link] and to others that are part of the following text material.

Applications of Electromagnetic Induction

There are many applications of Faraday’s Law of induction, as we will explore in this chapter and others. At this juncture, let us mention several that have to do with data storage and magnetic fields. A very important application has to do with audio and video *recording tapes*. A plastic tape, coated with iron oxide, moves past a recording head. This recording head is basically a round iron ring about which is wrapped a coil of wire—an electromagnet ([link]). A signal in the form of a varying input current from a microphone or camera goes to the recording head. These signals (which are a function of the signal amplitude and frequency) produce varying magnetic fields at the recording head. As the tape moves past the recording head, the magnetic field orientations of the iron oxide molecules on the tape are changed thus recording the signal. In the playback mode, the magnetized tape is run past another head, similar in structure to the recording head. The different magnetic field orientations of the iron oxide molecules on the tape induces an emf in the coil of wire in the playback head. This signal then is sent to a loudspeaker or video player.



Recording and playback heads used with audio and video magnetic tapes. (credit: Steve Jurvetson)

Similar principles apply to computer hard drives, except at a much faster rate. Here recordings are on a coated, spinning disk. Read heads historically were made to work on the principle of induction. However, the input information is carried in digital rather than analog form – a series of 0's or 1's are written upon the spinning hard drive. Today, most hard drive readout devices do not work on the principle of induction, but use a technique known as *giant magnetoresistance*. (The discovery that weak changes in a magnetic field in a thin film of iron and chromium could bring about much larger changes in electrical resistance was one of the first large successes of nanotechnology.) Another application of induction is found on the magnetic stripe on the back of your personal credit card as used at the grocery store or the ATM machine. This works on the same principle as the audio or video tape mentioned in the last paragraph in which a head reads personal information from your card.

Another application of electromagnetic induction is when electrical signals need to be transmitted across a barrier. Consider the *cochlear implant* shown below. Sound is picked up by a microphone on the outside of the skull and is used to set up a varying magnetic field. A current is induced in a receiver secured in the bone beneath the skin and transmitted to electrodes in the inner ear. Electromagnetic induction can be used in other instances where electric signals need to be conveyed across various media.

Electromagnetic induction used in transmitting electric currents across mediums. The device on the baby's head induces an electrical current in a receiver secured in the bone beneath the skin. (credit: Bjorn Knetsch)

Another contemporary area of research in which electromagnetic induction is being successfully implemented (and with substantial potential) is transcranial magnetic simulation. A host of disorders, including depression and hallucinations can be traced to irregular localized electrical activity in the brain. In *transcranial magnetic stimulation*, a rapidly varying and very localized magnetic field is placed close to certain sites identified in the brain. Weak electric currents are induced in the identified sites and can result in recovery of electrical functioning in the brain tissue.

Sleep apnea ("the cessation of breath") affects both adults and infants (especially premature babies and it may be a cause of sudden infant deaths [SID]). In such individuals, breath can stop repeatedly during their sleep. A cessation of more than 20 seconds can be very dangerous. Stroke, heart failure, and tiredness are just some of the possible consequences for a person having sleep apnea. The concern in infants is the stopping of breath for these longer times. One type of monitor to alert parents when a child is not breathing uses electromagnetic induction. A wire wrapped around the infant’s chest has an alternating current running through it. The expansion and contraction of the infant’s chest as the infant breathes changes the area through the coil. A pickup coil located nearby has an alternating current induced in it due to the changing magnetic field of the initial wire. If the child stops breathing, there will be a change in the induced current, and so a parent can be alerted.

Note:

Making Connections: Conservation of Energy

Lenz’s law is a manifestation of the conservation of energy. The induced emf produces a current that opposes the change in flux, because a change in flux means a change in energy. Energy can enter or leave, but not instantaneously. Lenz’s law is a consequence. As the change begins, the law says induction opposes and, thus, slows the change. In fact, if the induced emf were in the same direction as the change in flux, there would be a positive feedback that would give us free energy from no apparent source—conservation of energy would be violated.

Example:

Calculating Emf: How Great Is the Induced Emf?

Calculate the magnitude of the induced emf when the magnet in [link](a) is thrust into the coil, given the following information: the single loop coil has a radius of 6.00 cm and the average value of $B\cos\theta$ (this is given, since the bar magnet’s field is complex) increases from 0.0500 T to 0.250 T in 0.100 s.

Strategy

To find the *magnitude* of emf, we use Faraday’s law of induction as stated by $\text{emf} = -N\frac{\Delta\Phi}{\Delta t}$, but without the minus sign that indicates direction:

Equation:

$$\text{emf} = N\frac{\Delta\Phi}{\Delta t}.$$

Solution

We are given that $N = 1$ and $\Delta t = 0.100\text{ s}$, but we must determine the change in flux $\Delta\Phi$ before we can find emf. Since the area of the loop is fixed, we see that

Equation:

$$\Delta\Phi = \Delta(BA\ \cos\theta) = A\Delta(B\cos\theta\ ).$$

Now $\Delta(B\cos\theta) = 0.200\text{ T}$, since it was given that $B\cos\theta$ changes from 0.0500 to 0.250 T. The area of the loop is $A = \pi r^2 = (3.14...)(0.060\text{ m})^2 = 1.13\times10^{-2}\text{ m}^2$. Thus,

Equation:

$$\Delta\Phi = (1.13\times10^{-2}\text{ m}^2)(0.200\text{ T}\ ).$$

Entering the determined values into the expression for emf gives

Equation:

$$\text{Emf} = N\frac{\Delta\Phi}{\Delta t} = \frac{(1.13\times10^{-2}\text{ m}^2)(0.200\text{ T})}{0.100\text{ s}} = 22.6\text{ mV}.$$

Discussion

While this is an easily measured voltage, it is certainly not large enough for most practical applications. More loops in the coil, a stronger magnet, and faster movement make induction the practical source of voltages that it is.

Note:

PhET Explorations: Faraday's Electromagnetic Lab

Play with a bar magnet and coils to learn about Faraday's law. Move a bar magnet near one or two coils to make a light bulb glow. View the magnetic field lines. A meter shows the direction and magnitude of the current. View the magnetic field lines or use a meter to show the direction and magnitude of the current. You can also play with electromagnets, generators and transformers!

https://archive.cnx.org/specials/70b14c10-ae73-11e5-8eb2-b7fbe0c5c7a4/faraday/#sim-bar-magnet

Section Summary

- Faraday’s law of induction states that the emfinduced by a change in magnetic flux is
  Equation:
  $$\text{emf} = -N\frac{\Delta \Phi}{\Delta t}$$
  when flux changes by $\Delta \Phi$ in a time Δt.
- If emf is induced in a coil, N is its number of turns.
- The minus sign means that the emf creates a current I and magnetic field B that *oppose the change in flux* $\Delta \Phi$ —this opposition is known as Lenz’s law.

Conceptual Questions

Exercise:

Problem:

A person who works with large magnets sometimes places her head inside a strong field. She reports feeling dizzy as she quickly turns her head. How might this be associated with induction?

Exercise:

Problem:

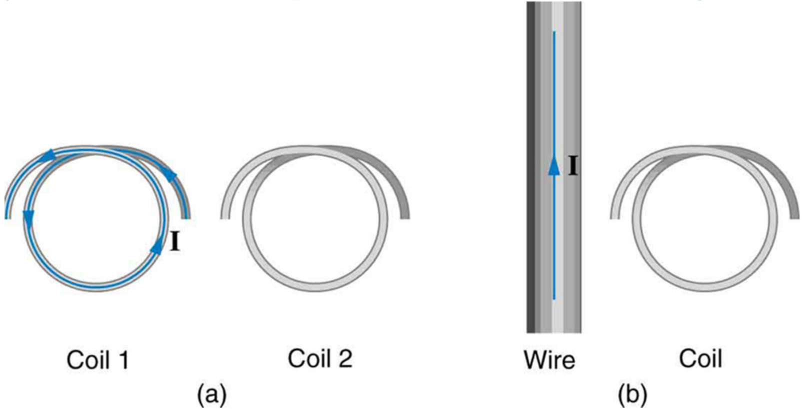
A particle accelerator sends high-velocity charged particles down an evacuated pipe. Explain how a coil of wire wrapped around the pipe could detect the passage of individual particles. Sketch a graph of the voltage output of the coil as a single particle passes through it.

Problems & Exercises

Exercise:

Problem:

Referring to [link](a), what is the direction of the current induced in coil 2: (a) If the current in coil 1 increases? (b) If the current in coil 1 decreases? (c) If the current in coil 1 is constant? Explicitly show how you follow the steps in the Problem-Solving Strategy for Lenz's Law.



(a) The coils lie in the same plane. (b) The wire is in the plane of the coil

---

Solution:

(a) CCW

(b) CW

(c) No current induced

Exercise:

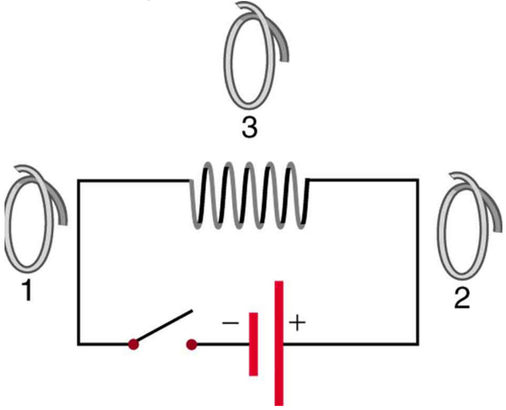
Problem:

Referring to [link](b), what is the direction of the current induced in the coil: (a) If the current in the wire increases? (b) If the current in the wire decreases? (c) If the current in the wire suddenly changes direction? Explicitly show how you follow the steps in the Problem-Solving Strategy for Lenz’s Law.

Exercise:

Problem:

Referring to [link], what are the directions of the currents in coils 1, 2, and 3 (assume that the coils are lying in the plane of the circuit): (a) When the switch is first closed? (b) When the switch has been closed for a long time? (c) Just after the switch is opened?



Solution:

(a) 1 CCW, 2 CCW, 3 CW

(b) 1, 2, and 3 no current induced

(c) 1 CW, 2 CW, 3 CCW

Exercise:

Problem: Repeat the previous problem with the battery reversed.

Exercise:

Problem:

Verify that the units of $\Delta\Phi/\ \Delta t$ are volts. That is, show that $1\ \text{T}\cdot\text{m}^2/\text{s} = 1\ \text{V}$.

Exercise:

Problem:

Suppose a 50-turn coil lies in the plane of the page in a uniform magnetic field that is directed into the page. The coil originally has an area of $0.250\ \text{m}^2$. It is stretched to have no area in 0.100 s. What is the direction and magnitude of the induced emf if the uniform magnetic field has a strength of 1.50 T?

Exercise:

Problem:

(a) An MRI technician moves his hand from a region of very low magnetic field strength into an MRI scanner's 2.00 T field with his fingers pointing in the direction of the field. Find the average emf induced in his wedding ring, given its diameter is 2.20 cm and assuming it takes 0.250 s to move it into the field. (b) Discuss whether this current would significantly change the temperature of the ring.

Solution:

(a) 3.04 mV
(b) As a lower limit on the ring, estimate R = 1.00 mΩ. The heat transferred will be 2.31 mJ. This is not a significant amount of heat.

Exercise:

Problem: Integrated Concepts

Referring to the situation in the previous problem: (a) What current is induced in the ring if its resistance is 0.0100 Ω? (b) What average power is dissipated? (c) What magnetic field is induced at the center of the ring? (d) What is the direction of the induced magnetic field relative to the MRI's field?

Exercise:

Problem:

An emf is induced by rotating a 1000-turn, 20.0 cm diameter coil in the Earth's 5.00×10^{-5} T magnetic field. What average emf is induced, given the plane of the coil is originally perpendicular to the Earth's field and is rotated to be parallel to the field in 10.0 ms?

Solution:

0.157 V

Exercise:

Problem:

A 0.250 m radius, 500-turn coil is rotated one-fourth of a revolution in 4.17 ms, originally having its plane perpendicular to a uniform magnetic field. (This is 60 rev/s.) Find the magnetic field strength needed to induce an average emf of 10,000 V.

Exercise:

Problem: Integrated Concepts

Approximately how does the emf induced in the loop in [link](b) depend on the distance of the center of the loop from the wire?

Solution:

proportional to $\frac{1}{r}$

Exercise:

Problem: **Integrated Concepts**

(a) A lightning bolt produces a rapidly varying magnetic field. If the bolt strikes the earth vertically and acts like a current in a long straight wire, it will induce a voltage in a loop aligned like that in [link](b). What voltage is induced in a 1.00 m diameter loop 50.0 m from a 2.00×10^6 A lightning strike, if the current falls to zero in 25.0 μs? (b) Discuss circumstances under which such a voltage would produce noticeable consequences.

Glossary

Faraday’s law of induction
: the means of calculating the emf in a coil due to changing magnetic flux, given by $\text{emf} = -N\frac{\Delta\Phi}{\Delta t}$

Lenz’s law
: the minus sign in Faraday’s law, signifying that the emf induced in a coil opposes the change in magnetic flux

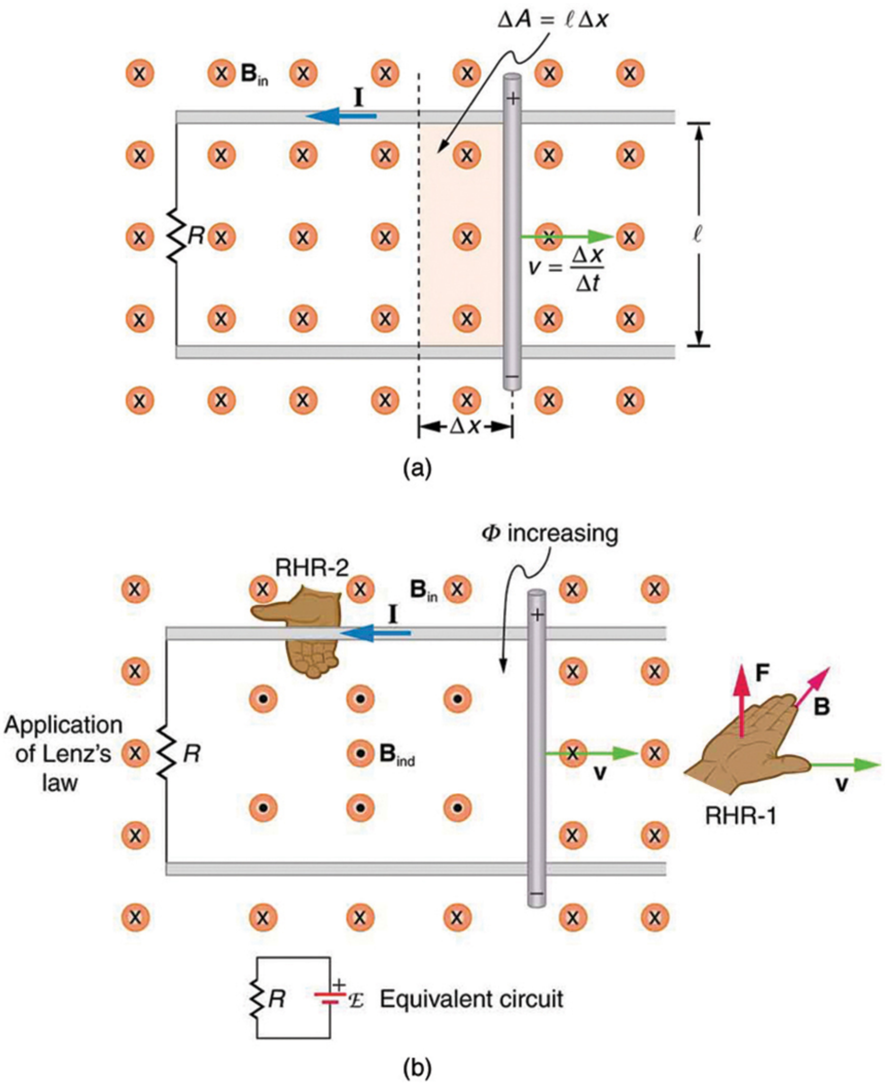
Motional Emf

- Calculate emf, force, magnetic field, and work due to the motion of an object in a magnetic field.

As we have seen, any change in magnetic flux induces an emf opposing that change—a process known as induction. Motion is one of the major causes of induction. For example, a magnet moved toward a coil induces an emf, and a coil moved toward a magnet produces a similar emf. In this section, we concentrate on motion in a magnetic field that is stationary relative to the Earth, producing what is loosely called *motional emf*.

One situation where motional emf occurs is known as the Hall effect and has already been examined. Charges moving in a magnetic field experience the magnetic force $F=\text{qvB}\,\text{sin}\,\theta$, which moves opposite charges in opposite directions and produces an $\text{emf}=B\ell v$. We saw that the Hall effect has applications, including measurements of B and v. We will now see that the Hall effect is one aspect of the broader phenomenon of induction, and we will find that motional emf can be used as a power source.

Consider the situation shown in [link]. A rod is moved at a speed v along a pair of conducting rails separated by a distance ℓ in a uniform magnetic field B. The rails are stationary relative to B and are connected to a stationary resistor R. The resistor could be anything from a light bulb to a voltmeter. Consider the area enclosed by the moving rod, rails, and resistor. B is perpendicular to this area, and the area is increasing as the rod moves. Thus the magnetic flux enclosed by the rails, rod, and resistor is increasing. When flux changes, an emf is induced according to Faraday’s law of induction.


ΔA = ℓΔx
B_in
I
R
ℓ
v = Δx/Δt
+
−
Δx
(a)
Φ increasing
RHR-2
B_in
I
Application of Lenz's law
R
B_ind
v
F
B
v
RHR-1
R
ℰ
Equivalent circuit
(b)

(a) A motional emf $= B\ell v$ is induced between the rails when this rod moves to the right in the uniform magnetic field. The magnetic field B is into the page, perpendicular to the moving rod and rails and, hence, to the area enclosed by them. (b) Lenz's law gives the directions of the induced field and current, and the polarity of the induced emf. Since the flux is increasing, the induced field is in the opposite direction, or out of the page. RHR-2 gives the current direction shown, and the polarity of the rod will drive such a current. RHR-1 also indicates the same polarity for the rod. (Note that the script E symbol used in the equivalent circuit at the bottom of part (b) represents emf.)

To find the magnitude of emf induced along the moving rod, we use Faraday’s law of induction without the sign:

Equation:

$$\text{emf} = N\frac{\Delta \Phi}{\Delta t}.$$

Here and below, “emf” implies the magnitude of the emf. In this equation, $N = 1$ and the flux $\Phi = \text{BA}\cos\theta$. We have $\theta = 0^{\text{o}}$ and $\cos\theta = 1$, since B is perpendicular to A. Now $\Delta\Phi = \Delta(\text{BA}) = B\Delta A$, since B is uniform. Note that the area swept out by the rod is $\Delta A = \ell\Delta x$. Entering these quantities into the expression for emf yields

Equation:

$$\text{emf} = \frac{B\Delta A}{\Delta t} = B\frac{\ell\Delta x}{\Delta t}.$$

Finally, note that $\Delta x/\Delta t = v$, the velocity of the rod. Entering this into the last expression shows that

Equation:

$$\text{emf} = B\ell v \qquad (B,\ \ell,\ \text{and}\ v\ \text{perpendicular})$$

is the motional emf. This is the same expression given for the Hall effect previously.

Note:

Making Connections: Unification of Forces

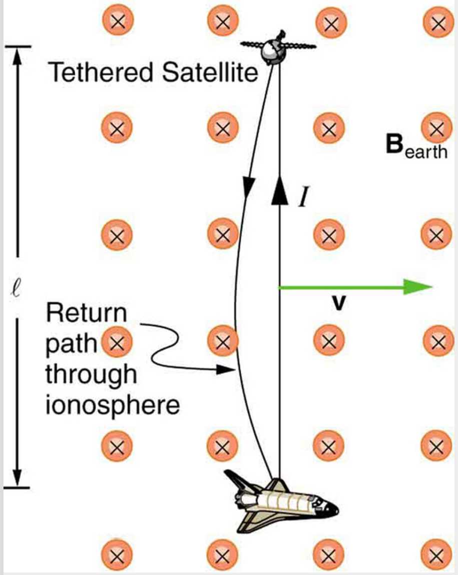
There are many connections between the electric force and the magnetic force. The fact that a moving electric field produces a magnetic field and, conversely, a moving magnetic field produces an electric field is part of why electric and magnetic forces are now considered to be different manifestations of the same force. This classic unification of electric and magnetic forces into what is called the electromagnetic force is the inspiration for contemporary efforts to unify other basic forces.

To find the direction of the induced field, the direction of the current, and the polarity of the induced emf, we apply Lenz's law as explained in Faraday's Law of Induction: Lenz's Law. (See [link](b).) Flux is increasing, since the area enclosed is increasing. Thus the induced field must oppose the existing one and be out of the page. And so the RHR-2 requires that *I* be counterclockwise, which in turn means the top of the rod is positive as shown.

Motional emf also occurs if the magnetic field moves and the rod (or other object) is stationary relative to the Earth (or some observer). We have seen an example of this in the situation where a moving magnet induces an emf in a stationary coil. It is the relative motion that is important. What is emerging in these observations is a connection between magnetic and electric fields. A moving magnetic field produces an electric field through its induced emf. We already have seen that a moving electric field produces a magnetic field—moving charge implies moving electric field and moving charge produces a magnetic field.

Motional emfs in the Earth's weak magnetic field are not ordinarily very large, or we would notice voltage along metal rods, such as a screwdriver, during ordinary motions. For example, a simple calculation of the motional emf of a 1 m rod moving at 3.0 m/s perpendicular to the Earth's field gives $\text{emf} = B\ell v = (5.0 \times 10^{-5}\ \text{T})(1.0\ \text{m})(3.0\ \text{m/s}) = 150\ \mu\text{V}$. This small value is consistent with experience. There is a spectacular exception, however. In 1992 and 1996, attempts were made with the space shuttle to create large motional emfs. The Tethered Satellite was to be let out on a 20 km length of wire as shown in [link], to create a 5 kV emf by moving at orbital speed through the Earth’s field. This emf could be used to convert some of the shuttle’s kinetic and potential energy into electrical energy if a complete circuit could be made. To complete the circuit, the stationary ionosphere was to supply a return path for the current to flow. (The ionosphere is the rarefied and partially ionized atmosphere at orbital altitudes. It conducts because of the ionization. The ionosphere serves the same function as the stationary rails and connecting resistor in [link], without which there would not be a complete circuit.) Drag on the current in the cable due to the magnetic force $F = I\ell B \sin \theta$ does the work that reduces the shuttle’s kinetic and potential energy and allows it to be converted to electrical energy. The tests were both unsuccessful. In the first, the cable hung up and could only be extended a couple of hundred meters; in the second, the cable broke when almost fully extended. [link] indicates feasibility in principle.

Example:
Calculating the Large Motional Emf of an Object in Orbit



Motional emf as electrical power conversion for the space shuttle is the motivation for the

Tethered Satellite experiment. A 5 kV emf was predicted to be induced in the 20 km long tether while moving at orbital speed in the Earth’s magnetic field. The circuit is completed by a return path through the stationary ionosphere.

Calculate the motional emf induced along a 20.0 km long conductor moving at an orbital speed of 7.80 km/s perpendicular to the Earth’s 5.00×10^{-5} T magnetic field.

Strategy

This is a straightforward application of the expression for motional emf—emf $= B\ell v$.

Solution

Entering the given values into emf $= B\ell v$ gives

Equation:

$$\begin{aligned} \text{emf} &= B\ell v \\ &= (5.00 \times 10^{-5}\ \text{T})(2.0 \times 10^{4}\ \text{m})(7.80 \times 10^{3}\ \text{m/s}) \\ &= 7.80 \times 10^{3}\ \text{V}. \end{aligned}$$

Discussion

The value obtained is greater than the 5 kV measured voltage for the shuttle experiment, since the actual orbital motion of the tether is not perpendicular to the Earth’s field. The 7.80 kV value is the maximum emf obtained when $\theta = 90^{\circ}$ and sin $\theta = 1$.

Section Summary

- An emf induced by motion relative to a magnetic field B is called a *motional emf* and is given by
  Equation:

  $$\text{emf} = B\ell v \qquad (B, \ell, \text{and } v \text{ perpendicular}),$$

  where ℓ is the length of the object moving at speed v relative to the field.

Conceptual Questions

Exercise:

Problem:

Why must part of the circuit be moving relative to other parts, to have usable motional emf? Consider, for example, that the rails in [link] are stationary relative to the magnetic field, while the rod moves.

Exercise:

Problem:

A powerful induction cannon can be made by placing a metal cylinder inside a solenoid coil. The cylinder is forcefully expelled when solenoid current is turned on rapidly. Use Faraday's and Lenz's laws to explain how this works. Why might the cylinder get live/hot when the cannon is fired?

Exercise:

Problem:

An induction stove heats a pot with a coil carrying an alternating current located beneath the pot (and without a hot surface). Can the stove surface be a conductor? Why won't a coil carrying a direct current work?

Exercise:

Problem:

Explain how you could thaw out a frozen water pipe by wrapping a coil carrying an alternating current around it. Does it matter whether or not the pipe is a conductor? Explain.

Problems & Exercises

Exercise:

Problem:

Use Faraday's law, Lenz's law, and RHR-1 to show that the magnetic force on the current in the moving rod in [link] is in the opposite direction of its velocity.

Exercise:

Problem:

If a current flows in the Satellite Tether shown in [link], use Faraday's law, Lenz's law, and RHR-1 to show that there is a magnetic force on the tether in the direction opposite to its velocity.

Exercise:

Problem:

(a) A jet airplane with a 75.0 m wingspan is flying at 280 m/s. What emf is induced between wing tips if the vertical component of the Earth's field is $3.00 \times 10^{-5}\ \text{T}$? (b) Is an emf of this magnitude likely to have any consequences? Explain.

Solution:

(a) 0.630 V

(b) No, this is a very small emf.

Exercise:

Problem:

(a) A nonferrous screwdriver is being used in a 2.00 T magnetic field. What maximum emf can be induced along its 12.0 cm length when it moves at 6.00 m/s? (b) Is it likely that this emf will have any consequences or even be noticed?

Exercise:

Problem:

At what speed must the sliding rod in [link] move to produce an emf of 1.00 V in a 1.50 T field, given the rod’s length is 30.0 cm?

Solution:

2.22 m/s

Exercise:

Problem:

The 12.0 cm long rod in [link] moves at 4.00 m/s. What is the strength of the magnetic field if a 95.0 V emf is induced?

Exercise:

Problem:

Prove that when B, ℓ, and v are not mutually perpendicular, motional emf is given by emf $= B\ell v\text{sin }\theta$. If v is perpendicular to B, then θ is the angle between ℓ and B. If ℓ is perpendicular to B, then θ is the angle between v and B.

Exercise:

Problem:

In the August 1992 space shuttle flight, only 250 m of the conducting tether considered in [link] could be let out. A 40.0 V motional emf was generated in the Earth's $5.00 \times 10^{-5}\ \text{T}$ field, while moving at $7.80 \times 10^{3}\ \text{m/s}$. What was the angle between the shuttle's velocity and the Earth's field, assuming the conductor was perpendicular to the field?

Exercise:

Problem: Integrated Concepts

Derive an expression for the current in a system like that in [link], under the following conditions. The resistance between the rails is R, the rails and the moving rod are identical in cross section A and have the same resistivity ρ. The distance between the rails is l, and the rod moves at constant speed v perpendicular to the uniform field B. At time zero, the moving rod is next to the resistance R.

Exercise:

Problem: Integrated Concepts

The Tethered Satellite in [link] has a mass of 525 kg and is at the end of a 20.0 km long, 2.50 mm diameter cable with the tensile strength of steel. (a) How much does the cable stretch if a 100 N force is exerted to pull the satellite in? (Assume the satellite and shuttle are at the same altitude above the Earth.) (b) What is the effective force constant of the cable? (c) How much energy is stored in it when stretched by the 100 N force?

Exercise:

Problem: Integrated Concepts

The Tethered Satellite discussed in this module is producing 5.00 kV, and a current of 10.0 A flows. (a) What magnetic drag force does this produce if the system is moving at 7.80 km/s? (b) How much kinetic energy is removed from the system in 1.00 h, neglecting any change in altitude or velocity during that time? (c) What is the change in velocity if the mass of the system is 100,000 kg? (d) Discuss the long term consequences (say, a week-long mission) on the space shuttle's orbit, noting what effect a decrease in velocity has and assessing the magnitude of the effect.

Solution:

(a) 10.0 N

(b) 2.81×10^8 J

(c) 0.36 m/s

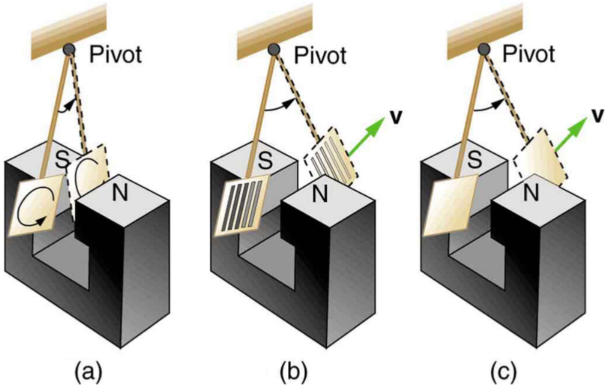
(d) For a week-long mission (168 hours), the change in velocity will be 60 m/s, or approximately 1%. In general, a decrease in velocity would cause the orbit to start spiraling inward because the velocity would no longer be sufficient to keep the circular orbit. The long-term consequences are that the shuttle would require a little more fuel to maintain the desired speed, otherwise the orbit would spiral slightly inward.

Eddy Currents and Magnetic Damping

- Explain the magnitude and direction of an induced eddy current, and the effect this will have on the object it is induced in.
- Describe several applications of magnetic damping.

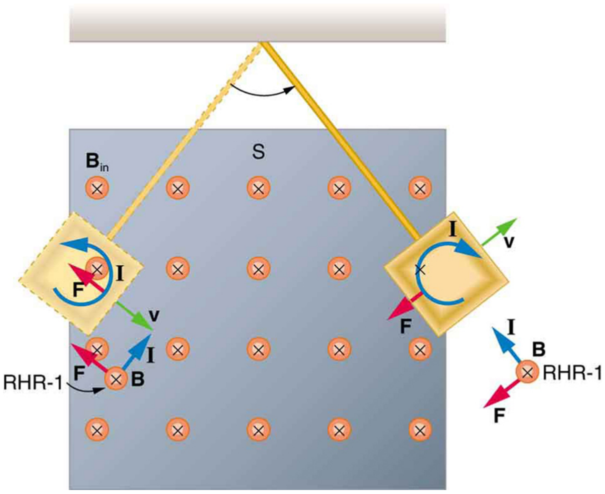
Eddy Currents and Magnetic Damping

As discussed in Motional Emf, motional emf is induced when a conductor moves in a magnetic field or when a magnetic field moves relative to a conductor. If motional emf can cause a current loop in the conductor, we refer to that current as an **eddy current**. Eddy currents can produce significant drag, called **magnetic damping**, on the motion involved. Consider the apparatus shown in [link], which swings a pendulum bob between the poles of a strong magnet. (This is another favorite physics lab activity.) If the bob is metal, there is significant drag on the bob as it enters and leaves the field, quickly damping the motion. If, however, the bob is a slotted metal plate, as shown in [link](b), there is a much smaller effect due to the magnet. There is no discernible effect on a bob made of an insulator. Why is there drag in both directions, and are there any uses for magnetic drag?



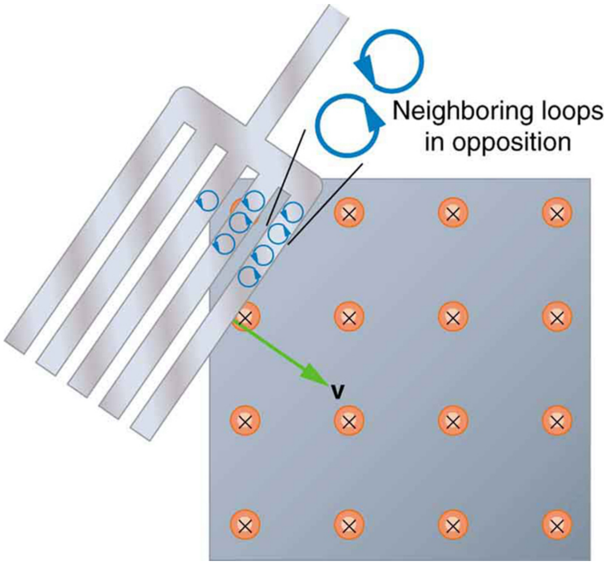
A common physics demonstration device for exploring eddy currents and magnetic damping. (a) The motion of a metal pendulum bob swinging between the poles of a magnet is quickly damped by the action of eddy currents. (b) There is little effect on the motion of a slotted metal bob, implying that eddy currents are made less effective. (c) There is also no magnetic damping on a nonconducting bob, since the eddy currents are extremely small.

[link] shows what happens to the metal plate as it enters and leaves the magnetic field. In both cases, it experiences a force opposing its motion. As it enters from the left, flux increases, and so an eddy current is set up (Faraday’s law) in the counterclockwise direction (Lenz’s law), as shown. Only the right-hand side of the current loop is in the field, so that there is an unopposed force on it to the left (RHR-1). When the metal plate is completely inside the field, there is no eddy current if the field is uniform, since the flux remains constant in this region. But when the plate leaves the field on the right, flux decreases, causing an eddy current in the clockwise direction that, again, experiences a force to the left, further slowing the motion. A similar analysis of what happens when the plate swings from the right toward the left shows that its motion is also damped when entering and leaving the field.



A more detailed look at the conducting plate passing between the poles of a magnet. As it enters and leaves the field, the change in flux produces an eddy current. Magnetic force on the current loop opposes the motion. There is no current and no magnetic drag when the plate is completely inside the uniform field.

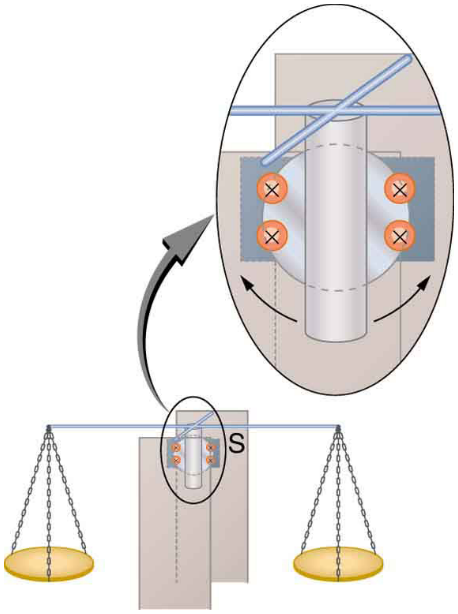
When a slotted metal plate enters the field, as shown in [link], an emf is induced by the change in flux, but it is less effective because the slots limit the size of the current loops. Moreover, adjacent loops have currents in opposite directions, and their effects cancel. When an insulating material is used, the eddy current is extremely small, and so magnetic damping on insulators is negligible. If eddy currents are to be avoided in conductors, then they can be slotted or constructed of thin layers of conducting material separated by insulating sheets.



Eddy currents induced in a slotted metal plate entering a magnetic field form small loops, and the forces on them tend to cancel, thereby making magnetic drag almost zero.

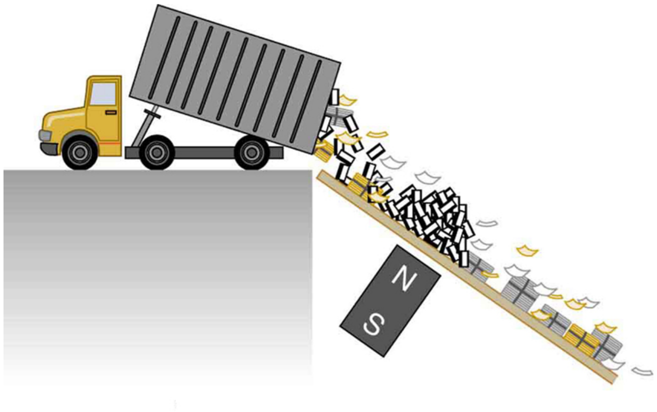
Applications of Magnetic Damping

One use of magnetic damping is found in sensitive laboratory balances. To have maximum sensitivity and accuracy, the balance must be as friction-free as possible. But if it is friction-free, then it will oscillate for a very long time. Magnetic damping is a simple and ideal solution. With magnetic damping, drag is proportional to speed and becomes zero at zero velocity. Thus the oscillations are quickly damped, after which the damping force disappears, allowing the balance to be very sensitive. (See [link].) In most balances, magnetic damping is accomplished with a conducting disc that rotates in a fixed field.



Magnetic damping of this sensitive balance slows its oscillations. Since Faraday’s law of induction gives the greatest effect for the most rapid change, damping is greatest for large oscillations and goes to zero as the motion stops.

Since eddy currents and magnetic damping occur only in conductors, recycling centers can use magnets to separate metals from other materials. Trash is dumped in batches down a ramp, beneath which lies a powerful magnet. Conductors in the trash are slowed by magnetic damping while nonmetals in the trash move on, separating from the metals. (See [link].) This works for all metals, not just ferromagnetic ones. A magnet can separate out the ferromagnetic materials alone by acting on stationary trash.



Metals can be separated from other trash by magnetic drag. Eddy currents and magnetic drag are created in the metals sent down this ramp by the powerful magnet beneath it. Nonmetals move on.

Other major applications of eddy currents are in metal detectors and braking systems in trains and roller coasters. Portable metal detectors ([link]) consist of a primary coil carrying an alternating current and a secondary coil in which a current is induced. An eddy current will be induced in a piece of metal close to the detector which will cause a change in the induced current within the secondary coil, leading to some sort of signal like a shrill noise. Braking using eddy currents is safer because factors such as rain do not affect the braking and the braking is smoother. However, eddy currents cannot bring the motion to a complete stop, since the force produced decreases with speed. Thus, speed can be reduced from say 20 m/s to 5 m/s, but another form of braking is needed to completely stop the vehicle. Generally, powerful rare earth magnets such as neodymium magnets are used in roller coasters. [link] shows rows of magnets in such an application. The vehicle has metal fins (normally containing copper) which pass through the magnetic field slowing the vehicle down in much the same way as with the pendulum bob shown in [link].

A soldier in Iraq uses a metal detector to search for explosives and weapons. (credit: U.S. Army)

The rows of rare earth magnets (protruding horizontally) are used for magnetic braking in roller coasters. (credit: Stefan Scheer, Wikimedia Commons)

Induction cooktops have electromagnets under their surface. The magnetic field is varied rapidly producing eddy currents in the base of the pot, causing the pot and its contents to increase in temperature. Induction cooktops have high efficiencies and good response times but the base of the pot needs to be ferromagnetic, iron or steel for induction to work.

Section Summary

- Current loops induced in moving conductors are called eddy currents.
- They can create significant drag, called magnetic damping.

Conceptual Questions

Exercise:

Problem:

Explain why magnetic damping might not be effective on an object made of several thin conducting layers separated by insulation.

Exercise:

Problem:

Explain how electromagnetic induction can be used to detect metals? This technique is particularly important in detecting buried landmines for disposal, geophysical prospecting and at airports.

Problems & Exercises

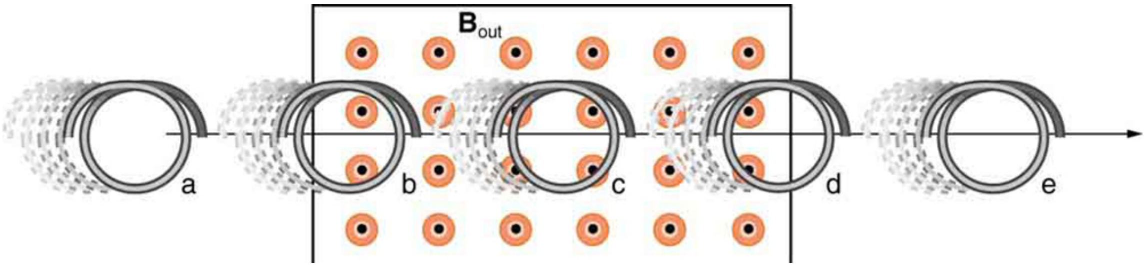
Exercise:

Problem:

Make a drawing similar to [link], but with the pendulum moving in the opposite direction. Then use Faraday’s law, Lenz’s law, and RHR-1 to show that magnetic force opposes motion.

Exercise:

Problem:



A coil is moved into and out of a region of uniform magnetic field.

A coil is moved through a magnetic field as shown in [link]. The field is uniform inside the rectangle and zero outside. What is the direction of the induced current and what is the direction of the magnetic force on the coil at each position shown?

Glossary

eddy current
a current loop in a conductor caused by motional emf

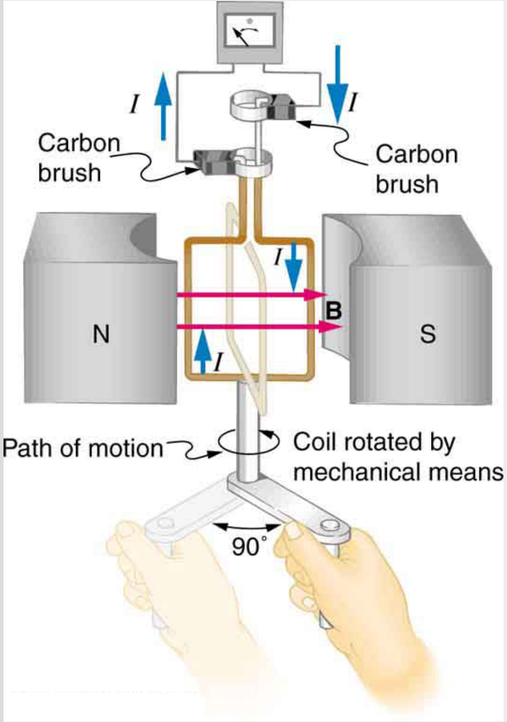
magnetic damping
the drag produced by eddy currents

Electric Generators

- Calculate the emf induced in a generator.
- Calculate the peak emf which can be induced in a particular generator system.

Electric generators induce an emf by rotating a coil in a magnetic field, as briefly discussed in Induced Emf and Magnetic Flux. We will now explore generators in more detail. Consider the following example.

Example:
Calculating the Emf Induced in a Generator Coil
The generator coil shown in [link] is rotated through one-fourth of a revolution (from $\theta = 0^\circ$ to $\theta = 90^\circ$) in 15.0 ms. The 200-turn circular coil has a 5.00 cm radius and is in a uniform 1.25 T magnetic field. What is the average emf induced?


I
I
Carbon brush
Carbon brush
I
B
N
S
I
Path of motion
Coil rotated by mechanical means
90°

When this generator coil is rotated through one-fourth of a revolution, the magnetic flux Φ changes from its maximum to zero, inducing an emf.

Strategy

We use Faraday’s law of induction to find the average emf induced over a time Δt:

Equation:

$$\text{emf} = -N\frac{\Delta\Phi}{\Delta t}.$$

We know that $N = 200$ and $\Delta t = 15.0\text{ ms}$, and so we must determine the change in flux $\Delta\Phi$ to find emf.

Solution

Since the area of the loop and the magnetic field strength are constant, we see that

Equation:

$$\Delta\Phi = \Delta(BA\cos\theta) = AB\Delta(\cos\theta).$$

Now, $\Delta(\cos\theta) = -1.0$, since it was given that θ goes from 0^{o} to 90^{o} . Thus $\Delta\Phi = -\text{AB}$, and

Equation:

$$\text{emf} = N\frac{\text{AB}}{\Delta t}.$$

The area of the loop is $A = \pi r^2 = (3.14...)(0.0500\text{ m})^2 = 7.85\times10^{-3}\text{ m}^2$. Entering this value gives

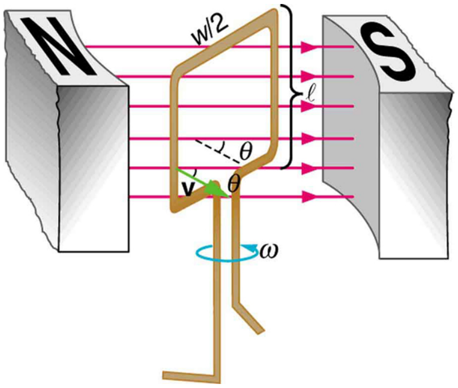
Equation:

$$\text{emf} = 200\frac{(7.85\times10^{-3}\text{ m}^2)(1.25\text{ T})}{15.0\times10^{-3}\text{ s}} = 131\text{ V}.$$

Discussion

This is a practical average value, similar to the 120 V used in household power.

The emf calculated in [link] is the average over one-fourth of a revolution. What is the emf at any given instant? It varies with the angle between the magnetic field and a perpendicular to the coil. We can get an expression for emf as a function of time by considering the motional emf on a rotating rectangular coil of width w and height ℓ in a uniform magnetic field, as illustrated in [link].



A generator with a single rectangular coil rotated at constant angular velocity in a uniform magnetic field produces an emf that varies sinusoidally in time. Note the generator is similar to a motor, except the shaft is rotated to produce a current rather than the other way around.

Charges in the wires of the loop experience the magnetic force, because they are moving in a magnetic field. Charges in the vertical wires experience forces parallel to the wire, causing currents. But those in the top and bottom segments feel a force perpendicular to the wire, which does not cause a current. We can thus find the induced emf by considering only the side wires. Motional emf is given to be $\text{emf} = B\ell v$, where the velocity v is perpendicular to the magnetic field B. Here the velocity is at an angle θ with B, so that its component perpendicular to B is $v \sin \theta$ (see [link]). Thus in this case the emf induced on each side is $\text{emf} = B\ell v \sin \theta$, and they are in the same direction. The total emf around the loop is then

Equation:

$$\text{emf} = 2B\ell v \sin \theta.$$

This expression is valid, but it does not give emf as a function of time. To find the time dependence of emf, we assume the coil rotates at a constant angular velocity ω. The angle θ is related to angular velocity by $\theta = \omega t$, so that

Equation:

$$\text{emf} = 2B\ell v \sin \omega t.$$

Now, linear velocity v is related to angular velocity ω by $v = r\omega$. Here $r = w/2$, so that $v = (w/2)\omega$, and

Equation:

$$\text{emf} = 2B\ell \frac{w}{2} \omega \sin \omega t = (\ell w) B\omega \sin \omega t.$$

Noting that the area of the loop is $A = \ell w$, and allowing for N loops, we find that

Equation:

$$\text{emf} = \text{NAB}\omega \sin \omega t$$

is the **emf induced in a generator coil** of N turns and area A rotating at a constant angular velocity ω in a uniform magnetic field B. This can also be expressed as

Equation:

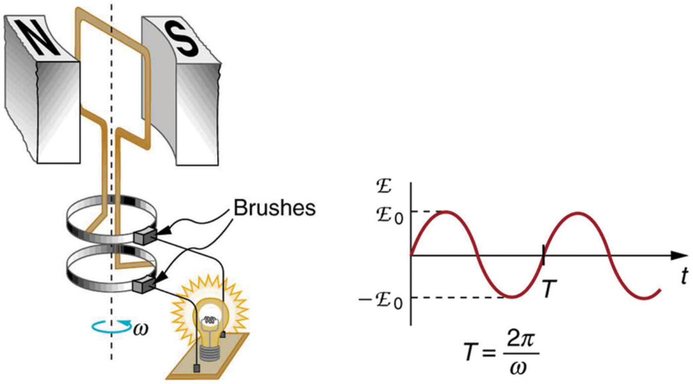
$$\text{emf} = \text{emf}_0 \sin \omega t,$$

where

Equation:

$$\text{emf}_0 = \text{NAB}\omega$$

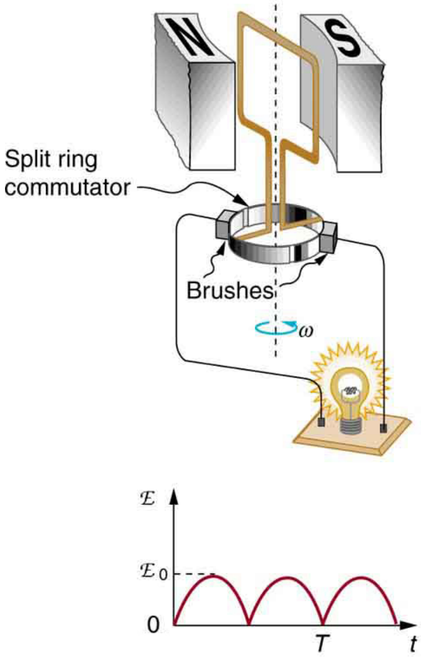
is the maximum **(peak) emf**. Note that the frequency of the oscillation is $f = \omega / 2\pi$, and the period is $T = 1/f = 2\pi/\omega$. [link] shows a graph of emf as a function of time, and it now seems reasonable that AC voltage is sinusoidal.



The emf of a generator is sent to a light bulb with the system of rings and brushes shown. The graph gives the emf of the generator as a function of time. emf_0 is the peak emf. The period is $T = 1/f = 2\pi/\omega$, where f is the frequency. Note that the script E stands for emf.

The fact that the peak emf, $\text{emf}_0 = \text{NAB}\omega$, makes good sense. The greater the number of coils, the larger their area, and the stronger the field, the greater the output voltage. It is interesting that the faster the generator is spun (greater ω), the greater the emf. This is noticeable on bicycle generators—at least the cheaper varieties. One of the authors as a juvenile found it amusing to ride his bicycle fast enough to burn out his lights, until he had to ride home lightless one dark night.

[link] shows a scheme by which a generator can be made to produce pulsed DC. More elaborate arrangements of multiple coils and split rings can produce smoother DC, although electronic rather than mechanical means are usually used to make ripple-free DC.



Split rings, called commutators, produce a pulsed DC emf output in this configuration.

Example:

Calculating the Maximum Emf of a Generator

Calculate the maximum emf, emf_0, of the generator that was the subject of [link].

Strategy

Once ω, the angular velocity, is determined, $\text{emf}_0 = \text{NAB}\omega$ can be used to find emf_0. All other quantities are known.

Solution

Angular velocity is defined to be the change in angle per unit time:

Equation:

$$\omega = \frac{\Delta\theta}{\Delta t}.$$

One-fourth of a revolution is $\pi/2$ radians, and the time is 0.0150 s; thus,

Equation:

$$\begin{array}{lll} \omega & = & \frac{\pi/2\text{ rad}}{0.0150\text{ s}} \\ & = & 104.7\text{ rad/s}. \end{array}$$

104.7 rad/s is exactly 1000 rpm. We substitute this value for ω and the information from the previous example into $\text{emf}_0 = \text{NAB}\omega$, yielding

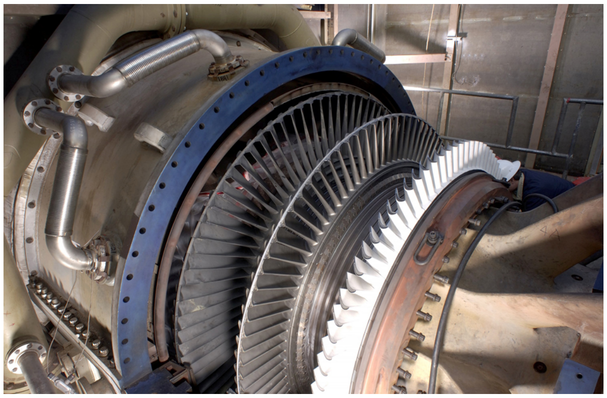
Equation:

$$\begin{array}{lll} \text{emf}_0 & = & NAB\omega \\ & = & 200(7.85\times 10^{-3}\text{ m}^2)(1.25\text{ T})(104.7\text{ rad/s}). \\ & = & 206\text{ V} \end{array}$$

Discussion

The maximum emf is greater than the average emf of 131 V found in the previous example, as it should be.

In real life, electric generators look a lot different than the figures in this section, but the principles are the same. The source of mechanical energy that turns the coil can be falling water (hydropower), steam produced by the burning of fossil fuels, or the kinetic energy of wind. [link] shows a cutaway view of a steam turbine; steam moves over the blades connected to the shaft, which rotates the coil within the generator.



Steam turbine/generator. The steam produced by burning coal impacts the turbine blades, turning the shaft which is connected to the generator. (credit: Nabonaco, Wikimedia Commons)

Generators illustrated in this section look very much like the motors illustrated previously. This is not coincidental. In fact, a motor becomes a generator when its shaft rotates. Certain early automobiles used their starter motor as a generator. In Back Emf, we shall further explore the action of a motor as a generator.

Section Summary

- An electric generator rotates a coil in a magnetic field, inducing an emfgiven as a function of time by
  Equation:

$$\text{emf} = \text{NAB}\omega \sin \omega t,$$

where A is the area of an N-turn coil rotated at a constant angular velocity ω in a uniform magnetic field B.

- The peak emf emf_0 of a generator is
  Equation:

$$\text{emf}_0 = \text{NAB}\omega.$$

Conceptual Questions

Exercise:

Problem:

Using RHR-1, show that the emfs in the sides of the generator loop in [link] are in the same sense and thus add.

Exercise:

Problem:

The source of a generator's electrical energy output is the work done to turn its coils. How is the work needed to turn the generator related to Lenz's law?

Problems & Exercises

Exercise:

Problem:

Calculate the peak voltage of a generator that rotates its 200-turn, 0.100 m diameter coil at 3600 rpm in a 0.800 T field.

Solution:

474 V

Exercise:

Problem:

At what angular velocity in rpm will the peak voltage of a generator be 480 V, if its 500-turn, 8.00 cm diameter coil rotates in a 0.250 T field?

Exercise:

Problem:

What is the peak emf generated by rotating a 1000-turn, 20.0 cm diameter coil in the Earth's $5.00 \times 10^{-5}\ \text{T}$ magnetic field, given the plane of the coil is originally perpendicular to the Earth's field and is rotated to be parallel to the field in 10.0 ms?

Solution:

0.247 V

Exercise:

Problem:

What is the peak emf generated by a 0.250 m radius, 500-turn coil is rotated one-fourth of a revolution in 4.17 ms, originally having its plane perpendicular to a uniform magnetic field. (This is 60 rev/s.)

Exercise:

Problem:

(a) A bicycle generator rotates at 1875 rad/s, producing an 18.0 V peak emf. It has a 1.00 by 3.00 cm rectangular coil in a 0.640 T field. How many turns are in the coil? (b) Is this number of turns of wire practical for a 1.00 by 3.00 cm coil?

Solution:

(a) 50

(b) yes

Exercise:

Problem: Integrated Concepts

This problem refers to the bicycle generator considered in the previous problem. It is driven by a 1.60 cm diameter wheel that rolls on the outside rim of the bicycle tire. (a) What is the velocity of the bicycle if the generator's angular velocity is 1875 rad/s? (b) What is the maximum emf of the generator when the bicycle moves at 10.0 m/s, noting that it was 18.0 V under the original conditions? (c) If the sophisticated generator can vary its own magnetic field, what field strength will it need at 5.00 m/s to produce a 9.00 V maximum emf?

Exercise:

Problem:

(a) A car generator turns at 400 rpm when the engine is idling. Its 300-turn, 5.00 by 8.00 cm rectangular coil rotates in an adjustable magnetic field so that it can produce sufficient voltage even at low rpms. What is the field strength needed to produce a 24.0 V peak emf? (b) Discuss how this required field strength compares to those available in permanent and electromagnets.

Solution:

(a) 0.477 T

(b) This field strength is small enough that it can be obtained using either a permanent magnet or an electromagnet.

Exercise:

Problem:

Show that if a coil rotates at an angular velocity ω, the period of its AC output is $2\pi/\omega$.

Exercise:

Problem:

A 75-turn, 10.0 cm diameter coil rotates at an angular velocity of 8.00 rad/s in a 1.25 T field, starting with the plane of the coil parallel to the field. (a) What is the peak emf? (b) At what time is the peak emf first reached? (c) At what time is the emf first at its most negative? (d) What is the period of the AC voltage output?

Solution:

(a) 5.89 V

(b) At t=0

(c) 0.393 s

(d) 0.785 s

Exercise:

Problem:

(a) If the emf of a coil rotating in a magnetic field is zero at $t=0$, and increases to its first peak at $t=0.100\text{ ms}$, what is the angular velocity of the coil? (b) At what time will its next maximum occur? (c) What is the period of the output? (d) When is the output first one-fourth of its maximum? (e) When is it next one-fourth of its maximum?

Exercise:

Problem: **Unreasonable Results**

A 500-turn coil with a 0.250 m^2 area is spun in the Earth's $5.00\times10^{-5}\text{ T}$ field, producing a 12.0 kV maximum emf. (a) At what angular velocity must the coil be spun? (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

(a) $1.92 \times 10^6\ \mathrm{rad/s}$

(b) This angular velocity is unreasonably high, higher than can be obtained for any mechanical system.

(c) The assumption that a voltage as great as 12.0 kV could be obtained is unreasonable.

Glossary

electric generator
: a device for converting mechanical work into electric energy; it induces an emf by rotating a coil in a magnetic field

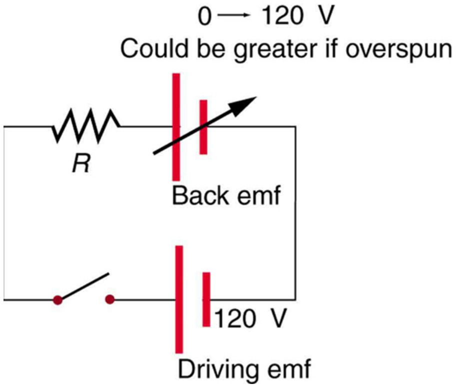
emf induced in a generator coil
: $\mathrm{emf} = \mathrm{NAB}\omega\ \sin\ \omega t$, where A is the area of an N-turn coil rotated at a constant angular velocity ω in a uniform magnetic field B, over a period of time t

peak emf
: $\mathrm{emf}_0 = \mathrm{NAB}\omega$

Back Emf

- Explain what back emf is and how it is induced.

It has been noted that motors and generators are very similar. Generators convert mechanical energy into electrical energy, whereas motors convert electrical energy into mechanical energy. Furthermore, motors and generators have the same construction. When the coil of a motor is turned, magnetic flux changes, and an emf (consistent with Faraday’s law of induction) is induced. The motor thus acts as a generator whenever its coil rotates. This will happen whether the shaft is turned by an external input, like a belt drive, or by the action of the motor itself. That is, when a motor is doing work and its shaft is turning, an emf is generated. Lenz’s law tells us the emf opposes any change, so that the input emf that powers the motor will be opposed by the motor’s self-generated emf, called the **back emf** of the motor. (See [link].)



The coil of a DC motor is represented as a resistor in this schematic. The back emf is represented as a variable emf that opposes the one driving the motor. Back emf is zero when the motor is not turning, and it increases

proportionally to the
motor’s angular
velocity.

Back emf is the generator output of a motor, and so it is proportional to the motor’s angular velocity ω. It is zero when the motor is first turned on, meaning that the coil receives the full driving voltage and the motor draws maximum current when it is on but not turning. As the motor turns faster and faster, the back emf grows, always opposing the driving emf, and reduces the voltage across the coil and the amount of current it draws. This effect is noticeable in a number of situations. When a vacuum cleaner, refrigerator, or washing machine is first turned on, lights in the same circuit dim briefly due to the IR drop produced in feeder lines by the large current drawn by the motor. When a motor first comes on, it draws more current than when it runs at its normal operating speed. When a mechanical load is placed on the motor, like an electric wheelchair going up a hill, the motor slows, the back emf drops, more current flows, and more work can be done. If the motor runs at too low a speed, the larger current can overheat it (via resistive power in the coil, $P = I^2R$), perhaps even burning it out. On the other hand, if there is no mechanical load on the motor, it will increase its angular velocity ω until the back emf is nearly equal to the driving emf. Then the motor uses only enough energy to overcome friction.

Consider, for example, the motor coils represented in [link]. The coils have a $0.400\ \Omega$ equivalent resistance and are driven by a 48.0 V emf. Shortly after being turned on, they draw a current $I = \mathrm{V/R} = (48.0\ \mathrm{V}\ )/\ (0.400\ \Omega) = 120\ \mathrm{A}$ and, thus, dissipate $P = I^2R = 5.76\ \mathrm{kW}$ of energy as heat transfer. Under normal operating conditions for this motor, suppose the back emf is 40.0 V. Then at operating speed, the total voltage across the coils is 8.0 V (48.0 V minus the 40.0 V back emf), and the current drawn is $I = \mathrm{V/R} = (8.0\ \mathrm{V})\ /(\ 0.400\ \Omega) = 20\ \mathrm{A}$. Under normal load, then, the power dissipated is $P = \mathrm{IV} = (20\ \mathrm{A})/(8.0\ \mathrm{V}) = 160\ \mathrm{W}$. The latter will not cause a problem for this motor, whereas the former 5.76 kW would burn out the coils if sustained.

Section Summary

- Any rotating coil will have an induced emf—in motors, this is called back emf, since it opposes the emf input to the motor.

Conceptual Questions

Exercise:

Problem:

Suppose you find that the belt drive connecting a powerful motor to an air conditioning unit is broken and the motor is running freely. Should you be worried that the motor is consuming a great deal of energy for no useful purpose? Explain why or why not.

Problems & Exercises

Exercise:

Problem:

Suppose a motor connected to a 120 V source draws 10.0 A when it first starts. (a) What is its resistance? (b) What current does it draw at its normal operating speed when it develops a 100 V back emf?

Solution:

(a) $12.00\ \Omega$

(b) 1.67 A

Exercise:

Problem:

A motor operating on 240 V electricity has a 180 V back emf at operating speed and draws a 12.0 A current. (a) What is its resistance? (b) What current does it draw when it is first started?

Exercise:

Problem:

What is the back emf of a 120 V motor that draws 8.00 A at its normal speed and 20.0 A when first starting?

Solution:

72.0 V

Exercise:

Problem:

The motor in a toy car operates on 6.00 V, developing a 4.50 V back emf at normal speed. If it draws 3.00 A at normal speed, what current does it draw when starting?

Exercise:

Problem: **Integrated Concepts**

The motor in a toy car is powered by four batteries in series, which produce a total emf of 6.00 V. The motor draws 3.00 A and develops a 4.50 V back emf at normal speed. Each battery has a $0.100\ \Omega$ internal resistance. What is the resistance of the motor?

Solution:

$0.100\ \Omega$

Glossary

back emf	the emf generated by a running motor, because it consists of a coil turning in a magnetic field; it opposes the voltage powering the motor

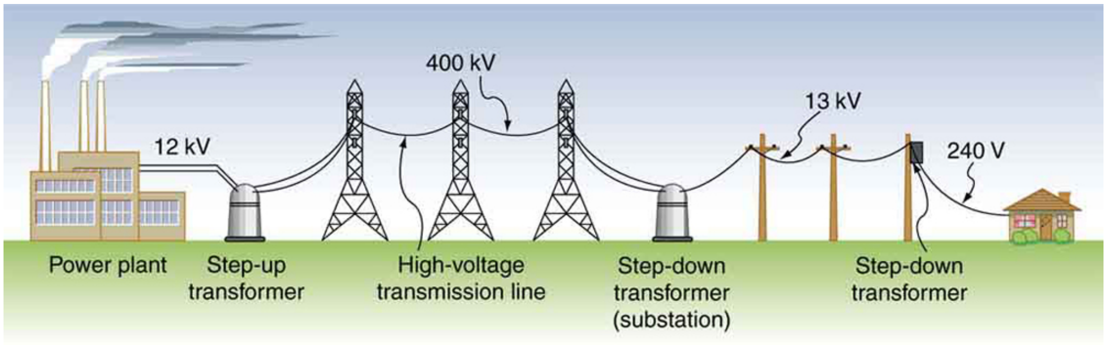
Transformers

- Explain how a transformer works.
- Calculate voltage, current, and/or number of turns given the other quantities.

Transformers do what their name implies—they transform voltages from one value to another (The term voltage is used rather than emf, because transformers have internal resistance). For example, many cell phones, laptops, video games, and power tools and small appliances have a transformer built into their plug-in unit (like that in [link]) that changes 120 V or 240 V AC into whatever voltage the device uses. Transformers are also used at several points in the power distribution systems, such as illustrated in [link]. Power is sent long distances at high voltages, because less current is required for a given amount of power, and this means less line loss, as was discussed previously. But high voltages pose greater hazards, so that transformers are employed to produce lower voltage at the user's location.

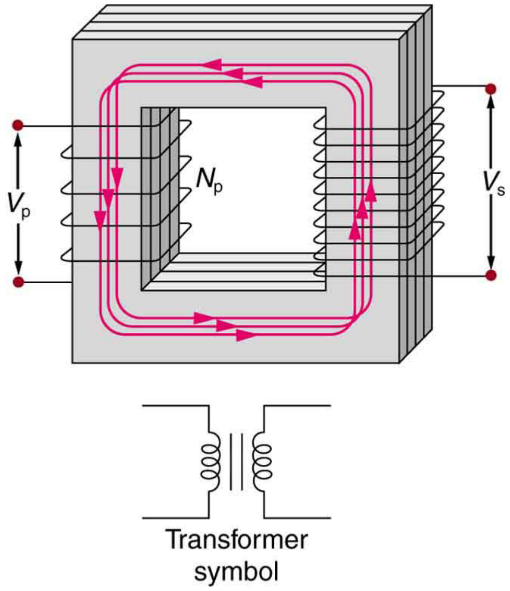
The plug-in transformer has become increasingly familiar with the proliferation of electronic devices that operate on voltages other than common 120 V

AC. Most are in the 3 to 12 V range. (credit: Shop Xtreme)


400 kV
13 kV
12 kV
240 V
Power plant
Step-up transformer
High-voltage transmission line
Step-down transformer (substation)
Step-down transformer

Transformers change voltages at several points in a power distribution system. Electric power is usually generated at greater than 10 kV, and transmitted long distances at voltages over 200 kV—sometimes as great as 700 kV—to limit energy losses. Local power distribution to neighborhoods or industries goes through a substation and is sent short distances at voltages ranging from 5 to 13 kV. This is reduced to 120, 240, or 480 V for safety at the individual user site.

The type of transformer considered in this text—see [link]—is based on Faraday’s law of induction and is very similar in construction to the apparatus Faraday used to demonstrate magnetic fields could cause currents. The two coils are called the *primary* and *secondary coils*. In normal use, the input voltage is placed on the primary, and the secondary produces the transformed output voltage. Not only does the iron core trap the magnetic field created by the primary coil, its magnetization increases the field strength. Since the input voltage is AC, a time-varying magnetic flux is sent to the secondary, inducing its AC output voltage.



A typical construction of a simple transformer has two coils wound on a ferromagnetic core that is laminated to minimize eddy currents. The magnetic field created by the primary is mostly confined to and increased by the core, which transmits it to the secondary coil. Any change in current in the primary induces a current in the secondary.

For the simple transformer shown in [link], the output voltage V_s depends almost entirely on the input voltage V_p and the ratio of the number of loops in the primary and secondary coils. Faraday's law of induction for the secondary coil gives its induced output voltage V_s to be

Equation:

$$V_\mathrm{s} = -N_\mathrm{s}\frac{\Delta\Phi}{\Delta t},$$

where N_s is the number of loops in the secondary coil and $\Delta\Phi/\ \Delta t$ is the rate of change of magnetic flux. Note that the output voltage equals the induced emf ($V_\mathrm{s} = \mathrm{emf}_\mathrm{s}$), provided coil resistance is small (a reasonable assumption for transformers). The cross-sectional area of the coils is the same on either side, as is the magnetic field strength, and so $\Delta\Phi/\Delta t$ is the same on either side. The input primary voltage V_p is also related to changing flux by

Equation:

$$V_p = -N_\mathrm{p}\frac{\Delta\Phi}{\Delta t}.$$

The reason for this is a little more subtle. Lenz's law tells us that the primary coil opposes the change in flux caused by the input voltage V_p, hence the minus sign (This is an example of *self-inductance*, a topic to be explored in some detail in later sections). Assuming negligible coil resistance, Kirchhoff's loop rule tells us that the induced emf exactly equals the input voltage. Taking the ratio of these last two equations yields a useful relationship:

Equation:

$$\frac{V_\mathrm{s}}{V_\mathrm{p}} = \frac{N_\mathrm{s}}{N_\mathrm{p}}.$$

This is known as the **transformer equation**, and it simply states that the ratio of the secondary to primary voltages in a transformer equals the ratio of the number of loops in their coils.

The output voltage of a transformer can be less than, greater than, or equal to the input voltage, depending on the ratio of the number of loops in their coils. Some transformers even provide a variable output by allowing connection to be made at different points on the secondary coil. A **step-up**

transformer is one that increases voltage, whereas a **step-down transformer** decreases voltage. Assuming, as we have, that resistance is negligible, the electrical power output of a transformer equals its input. This is nearly true in practice—transformer efficiency often exceeds 99%. Equating the power input and output,

Equation:

$$P_{\text{p}} = I_{\text{p}}V_{\text{p}} = I_{\text{s}}V_{\text{s}} = P_{\text{s}}.$$

Rearranging terms gives

Equation:

$$\frac{V_{\text{s}}}{V_{\text{p}}} = \frac{I_{\text{p}}}{I_{\text{s}}}.$$

Combining this with $\frac{V_{\text{s}}}{V_{\text{p}}} = \frac{N_{\text{s}}}{N_{\text{p}}}$, we find that

Equation:

$$\frac{I_{\text{s}}}{I_{\text{p}}} = \frac{N_{\text{p}}}{N_{\text{s}}}$$

is the relationship between the output and input currents of a transformer. So if voltage increases, current decreases. Conversely, if voltage decreases, current increases.

Example:

Calculating Characteristics of a Step-Up Transformer

A portable x-ray unit has a step-up transformer, the 120 V input of which is transformed to the 100 kV output needed by the x-ray tube. The primary has 50 loops and draws a current of 10.00 A when in use. (a) What is the number of loops in the secondary? (b) Find the current output of the secondary.

Strategy and Solution for (a)

We solve $\frac{V_{\text{s}}}{V_{\text{p}}} = \frac{N_{\text{s}}}{N_{\text{p}}}$ for N_{s}, the number of loops in the secondary, and enter the known values. This gives

Equation:

$$\begin{aligned} N_{\text{s}} &= N_{\text{p}}\frac{V_{\text{s}}}{V_{\text{p}}} \\ &= (50)\frac{100{,}000\text{ V}}{120\text{ V}} = 4.17 \times 10^{4}. \end{aligned}$$

Discussion for (a)

A large number of loops in the secondary (compared with the primary) is required to produce such a large voltage. This would be true for neon sign transformers and those supplying high voltage inside TVs and CRTs.

Strategy and Solution for (b)

We can similarly find the output current of the secondary by solving $\frac{I_{\text{s}}}{I_{\text{p}}} = \frac{N_{\text{p}}}{N_{\text{s}}}$ for I_{s} and entering known values. This gives

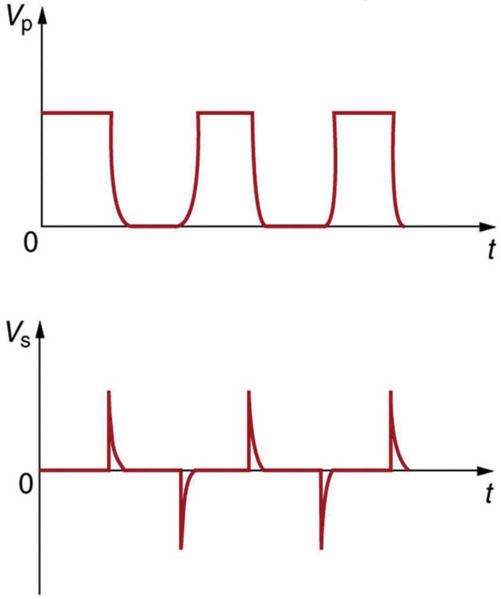
Equation:

$$\begin{aligned} I_{\text{s}} &= I_{\text{p}}\frac{N_{\text{p}}}{N_{\text{s}}} \\ &= (10.00\text{ A})\,\frac{50}{4.17\times 10^{4}} = 12.0\text{ mA}. \end{aligned}$$

Discussion for (b)

As expected, the current output is significantly less than the input. In certain spectacular demonstrations, very large voltages are used to produce long arcs, but they are relatively safe because the transformer output does not supply a large current. Note that the power input here is $P_{\text{p}} = I_{\text{p}}V_{\text{p}} = (10.00\text{ A})(120\text{ V}) = 1.20\text{ kW}$. This equals the power output $P_{\text{p}} = I_{\text{s}}V_{\text{s}} = (12.0\text{ mA})(100\text{ kV}) = 1.20\text{ kW}$, as we assumed in the derivation of the equations used.

The fact that transformers are based on Faraday’s law of induction makes it clear why we cannot use transformers to change DC voltages. If there is no change in primary voltage, there is no voltage induced in the secondary. One possibility is to connect DC to the primary coil through a switch. As the switch is opened and closed, the secondary produces a voltage like that in [link]. This is not really a practical alternative, and AC is in common use wherever it is necessary to increase or decrease voltages.



Transformers do not work for pure DC voltage input, but if it is switched on and off as on the top graph, the output will look something like that on the bottom graph. This is not the sinusoidal AC most AC appliances need.

Example:

Calculating Characteristics of a Step-Down Transformer

A battery charger meant for a series connection of ten nickel-cadmium batteries (total emf of 12.5 V DC) needs to have a 15.0 V output to charge the batteries. It uses a step-down transformer with a 200-loop primary and a 120 V input. (a) How many loops should there be in the secondary coil? (b) If the charging current is 16.0 A, what is the input current?

Strategy and Solution for (a)

You would expect the secondary to have a small number of loops. Solving $\frac{V_s}{V_p} = \frac{N_s}{N_p}$ for N_s and entering known values gives

Equation:

$$\begin{aligned} N_s &= N_p \frac{V_s}{V_p} \\ &= (200)\frac{15.0\text{ V}}{120\text{ V}} = 25. \end{aligned}$$

Strategy and Solution for (b)

The current input can be obtained by solving $\frac{I_s}{I_p} = \frac{N_p}{N_s}$ for I_p and entering known values. This gives

Equation:

$$\begin{aligned} I_p &= I_s \frac{N_s}{N_p} \\ &= (16.0\text{ A})\frac{25}{200} = 2.00\text{ A}. \end{aligned}$$

Discussion

The number of loops in the secondary is small, as expected for a step-down transformer. We also see that a small input current produces a larger output current in a step-down transformer. When transformers are used to operate large magnets, they sometimes have a small number of very heavy loops in the secondary. This allows the secondary to have low internal resistance and produce large currents. Note again that this solution is based on the assumption of 100% efficiency—or power out equals power in ($P_p = P_s$) —reasonable for good transformers. In this case the primary and secondary power is 240 W. (Verify this for yourself as a consistency check.) Note that the Ni-Cd batteries need to be charged from a DC power source (as would a 12 V battery). So the AC output of the secondary coil needs to be converted into DC. This is done using something called a rectifier, which uses devices called diodes that allow only a one-way flow of current.

Transformers have many applications in electrical safety systems, which are discussed in Electrical Safety: Systems and Devices.

Note:
PhET Explorations: Generator
Generate electricity with a bar magnet! Discover the physics behind the phenomena by exploring magnets and how you can use them to make a bulb light.
https://archive.cnx.org/specials/1e9b7292-ae74-11e5-a9dc-c7c8521ba8e6/generator/#sim-generator

Section Summary

- Transformers use induction to transform voltages from one value to another.
- For a transformer, the voltages across the primary and secondary coils are related by
  Equation:
  $$\frac{V_\text{s}}{V_\text{p}} = \frac{N_\text{s}}{N_\text{p}},$$
  where V_p and V_s are the voltages across primary and secondary coils having N_p and N_s turns.
- The currents I_p and I_s in the primary and secondary coils are related by $\frac{I_\text{s}}{I_\text{p}} = \frac{N_\text{p}}{N_\text{s}}$.
- A step-up transformer increases voltage and decreases current, whereas a step-down transformer decreases voltage and increases current.

Conceptual Questions

Exercise:

Problem:

Explain what causes physical vibrations in transformers at twice the frequency of the AC power involved.

Problems & Exercises

Exercise:

Problem:

A plug-in transformer, like that in [link], supplies 9.00 V to a video game system. (a) How many turns are in its secondary coil, if its input voltage is 120 V and the primary coil has 400 turns? (b) What is its input current when its output is 1.30 A?

Solution:

(a) 30.0

(b) 9.75×10^{-2} A

Exercise:

Problem:

An American traveler in New Zealand carries a transformer to convert New Zealand’s standard 240 V to 120 V so that she can use some small appliances on her trip. (a) What is the ratio of turns in the primary and secondary coils of her transformer? (b) What is the ratio of input to output current? (c) How could a New Zealander traveling in the United States use this same transformer to power her 240 V appliances from 120 V?

Exercise:

Problem:

A cassette recorder uses a plug-in transformer to convert 120 V to 12.0 V, with a maximum current output of 200 mA. (a) What is the current input? (b) What is the power input? (c) Is this amount of power reasonable for a small appliance?

Solution:

(a) 20.0 mA

(b) 2.40 W

(c) Yes, this amount of power is quite reasonable for a small appliance.

Exercise:

Problem:

(a) What is the voltage output of a transformer used for rechargeable flashlight batteries, if its primary has 500 turns, its secondary 4 turns, and the input voltage is 120 V? (b) What input current is required to produce a 4.00 A output? (c) What is the power input?

Exercise:

Problem:

(a) The plug-in transformer for a laptop computer puts out 7.50 V and can supply a maximum current of 2.00 A. What is the maximum input current if the input voltage is 240 V? Assume 100% efficiency. (b) If the actual efficiency is less than 100%, would the input current need to be greater or smaller? Explain.

Solution:

(a) 0.063 A

(b) Greater input current needed.

Exercise:

Problem:

A multipurpose transformer has a secondary coil with several points at which a voltage can be extracted, giving outputs of 5.60, 12.0, and 480 V. (a) The input voltage is 240 V to a primary coil of 280 turns. What are the numbers of turns in the parts of the secondary used to produce the output voltages? (b) If the maximum input current is 5.00 A, what are the maximum output currents (each used alone)?

Exercise:

Problem:

A large power plant generates electricity at 12.0 kV. Its old transformer once converted the voltage to 335 kV. The secondary of this transformer is being replaced so that its output can be 750 kV for more efficient cross-country transmission on upgraded transmission lines. (a) What is the ratio of turns in the new secondary compared with the old secondary? (b) What is the ratio of new current output to old output (at 335 kV) for the same power? (c) If the upgraded transmission lines have the same resistance, what is the ratio of new line power loss to old?

Solution:

(a) 2.2

(b) 0.45

(c) 0.20, or 20.0%

Exercise:

Problem:

If the power output in the previous problem is 1000 MW and line resistance is $2.00\ \Omega$, what were the old and new line losses?

Exercise:

Problem: **Unreasonable Results**

The 335 kV AC electricity from a power transmission line is fed into the primary coil of a transformer. The ratio of the number of turns in the secondary to the number in the primary is $N_{\text{s}}/N_{\text{p}} = 1000$. (a) What voltage is induced in the secondary? (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Solution:

(a) 335 MV

(b) way too high, well beyond the breakdown voltage of air over reasonable distances

(c) input voltage is too high

Exercise:

Problem: Construct Your Own Problem

Consider a double transformer to be used to create very large voltages. The device consists of two stages. The first is a transformer that produces a much larger output voltage than its input. The output of the first transformer is used as input to a second transformer that further increases the voltage. Construct a problem in which you calculate the output voltage of the final stage based on the input voltage of the first stage and the number of turns or loops in both parts of both transformers (four coils in all). Also calculate the maximum output current of the final stage based on the input current. Discuss the possibility of power losses in the devices and the effect on the output current and power.

Glossary

transformer
: a device that transforms voltages from one value to another using induction

transformer equation
: the equation showing that the ratio of the secondary to primary voltages in a transformer equals the ratio of the number of loops in their coils; $\frac{V_s}{V_p} = \frac{N_s}{N_p}$

step-up transformer
: a transformer that increases voltage

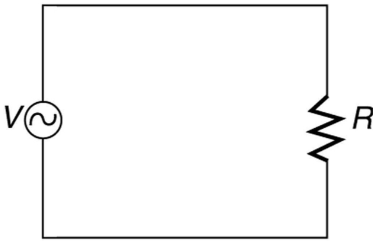
step-down transformer
a transformer that decreases voltage

Electrical Safety: Systems and Devices

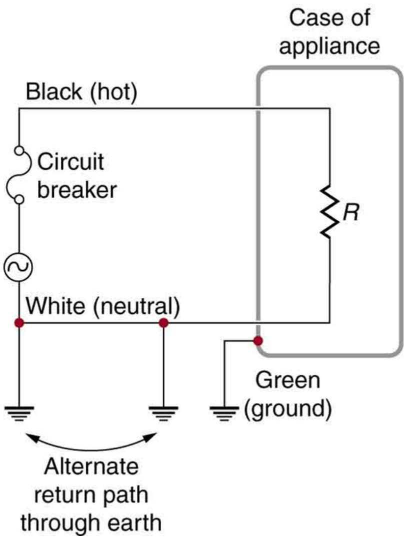
- Explain how various modern safety features in electric circuits work, with an emphasis on how induction is employed.

Electricity has two hazards. A **thermal hazard** occurs when there is electrical overheating. A **shock hazard** occurs when electric current passes through a person. Both hazards have already been discussed. Here we will concentrate on systems and devices that prevent electrical hazards.

[link] shows the schematic for a simple AC circuit with no safety features. This is not how power is distributed in practice. Modern household and industrial wiring requires the **three-wire system**, shown schematically in [link], which has several safety features. First is the familiar *circuit breaker* (or *fuse*) to prevent thermal overload. Second, there is a protective *case* around the appliance, such as a toaster or refrigerator. The case's safety feature is that it prevents a person from touching exposed wires and coming into electrical contact with the circuit, helping prevent shocks.

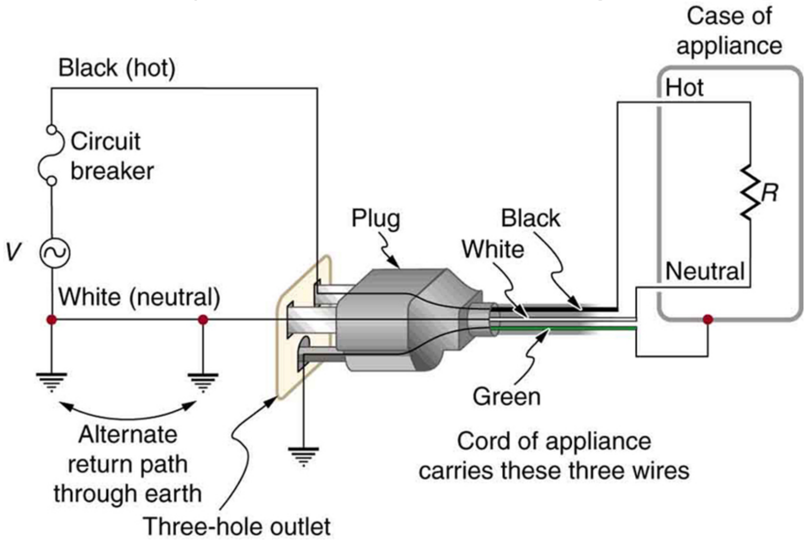


Schematic of a simple AC circuit with a voltage source and a single appliance represented by the resistance R. There are no safety features in this circuit.


Case of appliance
Black (hot)
Circuit breaker
R
White (neutral)
Green (ground)
Alternate return path through earth

The three-wire system connects the neutral wire to the earth at the voltage source and user location, forcing it to be at zero volts and supplying an alternative return path for the current through the earth. Also grounded to zero volts is the case of the appliance. A circuit breaker or fuse protects against thermal overload and is in series on the active (live/hot) wire. Note that wire insulation colors vary with region and it is essential to check locally to determine which color codes are in use (and even if they were followed in the particular installation).

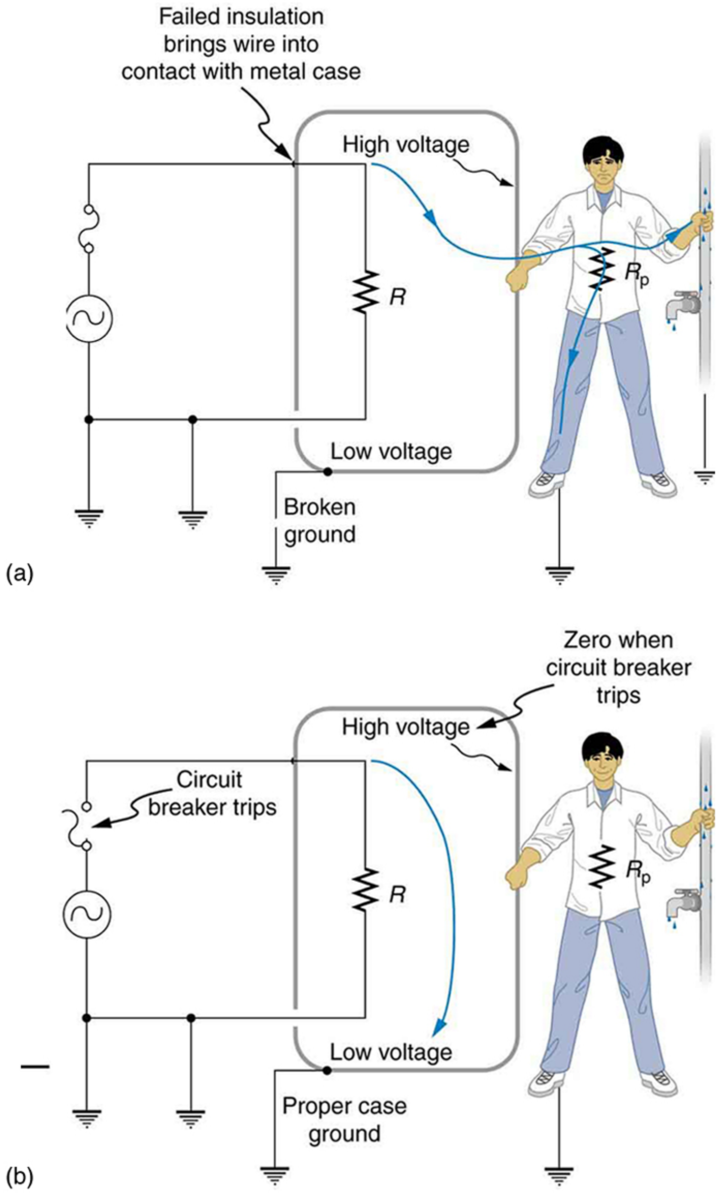
There are *three connections to earth or ground* (hereafter referred to as “earth/ground”) shown in [link]. Recall that an earth/ground connection is a low-resistance path directly to the earth. The two earth/ground connections on the *neutral wire* force it to be at zero volts relative to the earth, giving the wire its name. This wire is therefore safe to touch even if its insulation, usually white, is missing. The neutral wire is the return path for the current to follow to complete the circuit. Furthermore, the two earth/ground connections supply an alternative path through the earth, a good conductor, to complete the circuit. The earth/ground connection closest to the power source could be at the generating plant, while the other is at the user’s location. The third earth/ground is to the case of the appliance, through the green *earth/ground wire*, forcing the case, too, to be at zero volts. The *live* or *hot wire* (hereafter referred to as “live/hot”) supplies voltage and current to operate the appliance. [link] shows a more pictorial version of how the three-wire system is connected through a three-prong plug to an appliance.


Case of appliance
Black (hot)
Hot
Circuit breaker
R
Plug
Black
White
V
Neutral
White (neutral)
Green
Alternate return path through earth
Cord of appliance carries these three wires
Three-hole outlet

The standard three-prong plug can only be inserted in one way, to assure proper function of the three-wire system.

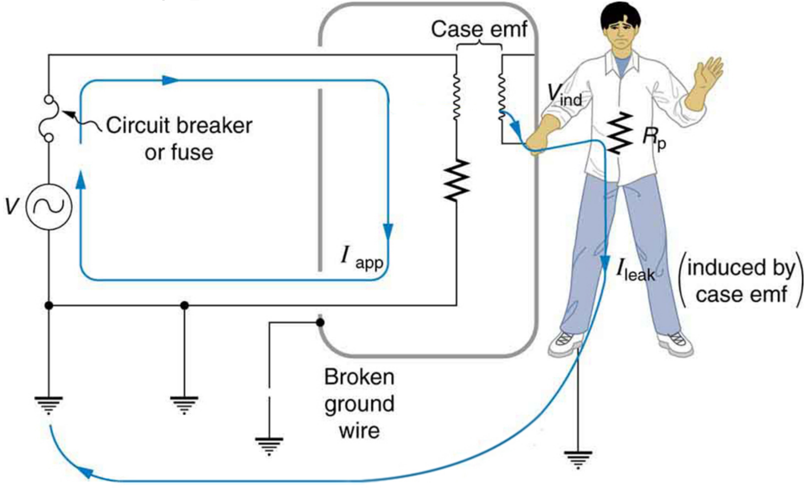
A note on insulation color-coding: Insulating plastic is color-coded to identify live/hot, neutral and ground wires but these codes vary around the world. Live/hot wires may be brown, red, black, blue or grey. Neutral wire may be blue, black or white. Since the same color may be used for live/hot or neutral in different parts of the world, it is essential to determine the color code in your region. The only exception is the earth/ground wire which is often green but may be yellow or just bare wire. Striped coatings are sometimes used for the benefit of those who are colorblind.

The three-wire system replaced the older two-wire system, which lacks an earth/ground wire. Under ordinary circumstances, insulation on the live/hot and neutral wires prevents the case from being directly in the circuit, so that the earth/ground wire may seem like double protection. Grounding the case solves more than one problem, however. The simplest problem is worn insulation on the live/hot wire that allows it to contact the case, as shown in [link]. Lacking an earth/ground connection (some people cut the third prong off the plug because they only have outdated two hole receptacles), a severe shock is possible. This is particularly dangerous in the kitchen, where a good connection to earth/ground is available through water on the floor or a water faucet. With the earth/ground connection intact, the circuit breaker will trip, forcing repair of the appliance. Why are some appliances still sold with two-prong plugs? These have nonconducting cases, such as power tools with impact resistant plastic cases, and are called *doubly insulated*. Modern two-prong plugs can be inserted into the asymmetric standard outlet in only one way, to ensure proper connection of live/hot and neutral wires.


Failed insulation brings wire into contact with metal case
High voltage
R
R_p
Low voltage
Broken ground
(a)
Zero when circuit breaker trips
High voltage
Circuit breaker trips
R
R_p
Low voltage
Proper case ground
(b)

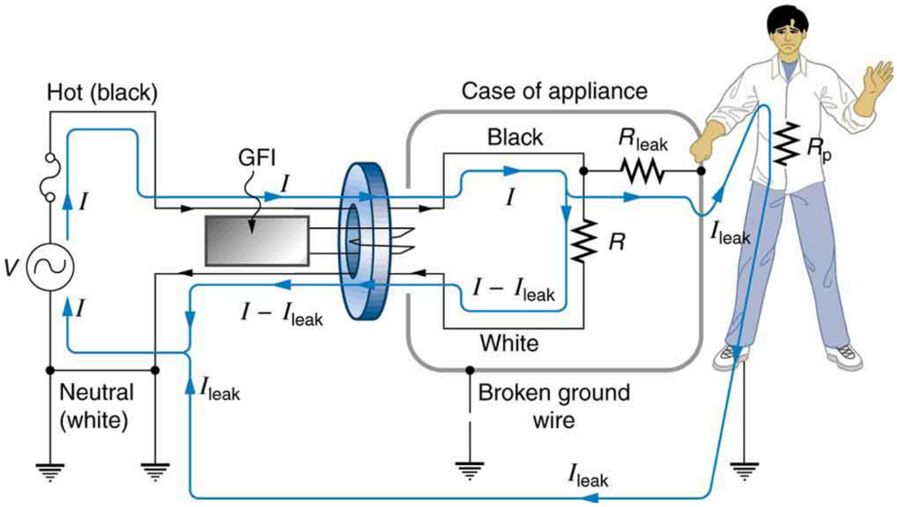
Worn insulation allows the live/hot wire to come into direct contact with the metal case of this appliance. (a) The earth/ground connection being broken, the person is severely shocked. The appliance may operate normally in this situation. (b) With a proper earth/ground, the circuit breaker trips, forcing repair of the appliance.

Electromagnetic induction causes a more subtle problem that is solved by grounding the case. The AC current in appliances can induce an emf on the case. If grounded, the case voltage is kept near zero, but if the case is not grounded, a shock can occur as pictured in [link]. Current driven by the induced case emf is called a *leakage current*, although current does not necessarily pass from the resistor to the case.



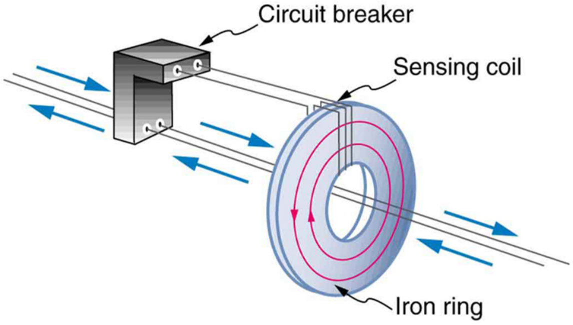
AC currents can induce an emf on the case of an appliance. The voltage can be large enough to cause a shock. If the case is grounded, the induced emf is kept near zero.

A *ground fault interrupter* (GFI) is a safety device found in updated kitchen and bathroom wiring that works based on electromagnetic induction. GFIs compare the currents in the live/hot and neutral wires. When live/hot and neutral currents are not equal, it is almost always because current in the neutral is less than in the live/hot wire. Then some of the current, again called a leakage current, is returning to the voltage source by a path other than through the neutral wire. It is assumed that this path presents a hazard, such as shown in [link]. GFIs are usually set to interrupt the circuit if the leakage current is greater than 5 mA, the accepted maximum harmless shock. Even if the leakage current goes safely to earth/ground through an intact earth/ground wire, the GFI will trip, forcing repair of the leakage.


Hot (black)
GFI
I
I
Case of appliance
Black
R_{leak}
R_p
I
R
I_{leak}
V
I
$I - I_{leak}$
$I - I_{leak}$
White
Neutral (white)
I_{leak}
Broken ground wire
I_{leak}

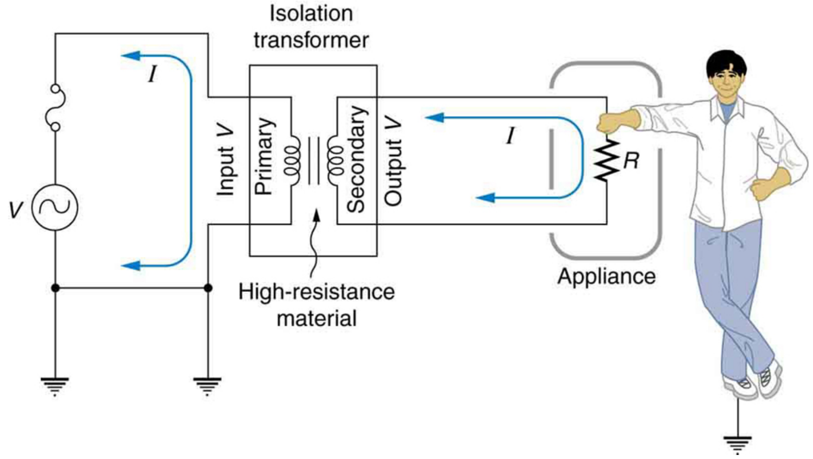
A ground fault interrupter (GFI) compares the currents in the live/hot and neutral wires and will trip if their difference exceeds a safe value. The leakage current here follows a hazardous path that could have been prevented by an intact earth/ground wire.

[link] shows how a GFI works. If the currents in the live/hot and neutral wires are equal, then they induce equal and opposite emfs in the coil. If not, then the circuit breaker will trip.


Circuit breaker
Sensing coil
Iron ring

A GFI compares currents by using both to induce an emf in the same coil. If the currents are equal, they will induce equal but opposite emfs.

Another induction-based safety device is the *isolation transformer*, shown in [link]. Most isolation transformers have equal input and output voltages. Their function is to put a large resistance between the original voltage source and the device being operated. This prevents a complete circuit between them, even in the circumstance shown. There is a complete circuit through the appliance. But there is not a complete circuit for current to flow through the person in the figure, who is touching only one of the transformer’s output wires, and neither output wire is grounded. The appliance is isolated from the original voltage source by the high resistance of the material between the transformer coils, hence the name isolation transformer. For current to flow through the person, it must pass through the high-resistance material between the coils, through the wire, the person, and back through the earth—a path with such a large resistance that the current is negligible.



An isolation transformer puts a large resistance between the original voltage source and the device, preventing a complete circuit between them.

The basics of electrical safety presented here help prevent many electrical hazards. Electrical safety can be pursued to greater depths. There are, for example, problems related to different earth/ground connections for appliances in close proximity. Many other examples are found in hospitals. Microshock-sensitive patients, for instance, require special protection. For these people, currents as low as 0.1 mA may cause ventricular fibrillation. The interested reader can use the material presented here as a basis for further study.

Section Summary

- Electrical safety systems and devices are employed to prevent thermal and shock hazards.
- Circuit breakers and fuses interrupt excessive currents to prevent thermal hazards.
- The three-wire system guards against thermal and shock hazards, utilizing live/hot, neutral, and earth/ground wires, and grounding the neutral wire and case of the appliance.
- A ground fault interrupter (GFI) prevents shock by detecting the loss of current to unintentional paths.
- An isolation transformer insulates the device being powered from the original source, also to prevent shock.
- Many of these devices use induction to perform their basic function.

Conceptual Questions

Exercise:

Problem:

Does plastic insulation on live/hot wires prevent shock hazards, thermal hazards, or both?

Exercise:

Problem:

Why are ordinary circuit breakers and fuses ineffective in preventing shocks?

Exercise:

Problem:

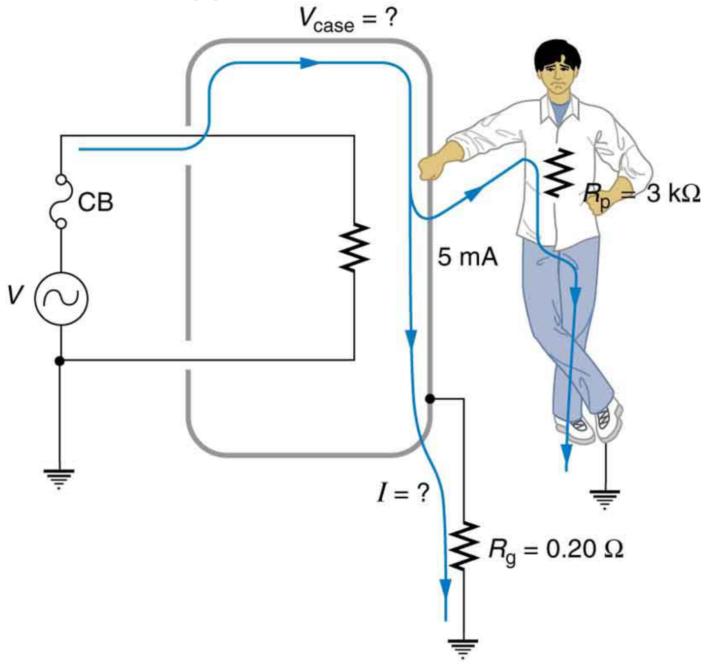
A GFI may trip just because the live/hot and neutral wires connected to it are significantly different in length. Explain why.

Problems & Exercises

Exercise:

Problem: **Integrated Concepts**

A short circuit to the grounded metal case of an appliance occurs as shown in [link]. The person touching the case is wet and only has a $3.00\ \text{k}\Omega$ resistance to earth/ground. (a) What is the voltage on the case if 5.00 mA flows through the person? (b) What is the current in the short circuit if the resistance of the earth/ground wire is $0.200\ \Omega$? (c) Will this trigger the 20.0 A circuit breaker supplying the appliance?



A person can be shocked even when the case of an appliance is grounded. The large short circuit current produces a voltage on the case of the appliance, since the resistance of the earth/ground wire is not zero.

Solution:

(a) 15.0 V

(b) 75.0 A

(c) yes

Glossary

thermal hazard
the term for electrical hazards due to overheating

shock hazard
the term for electrical hazards due to current passing through a human

three-wire system
the wiring system used at present for safety reasons, with live, neutral, and ground wires

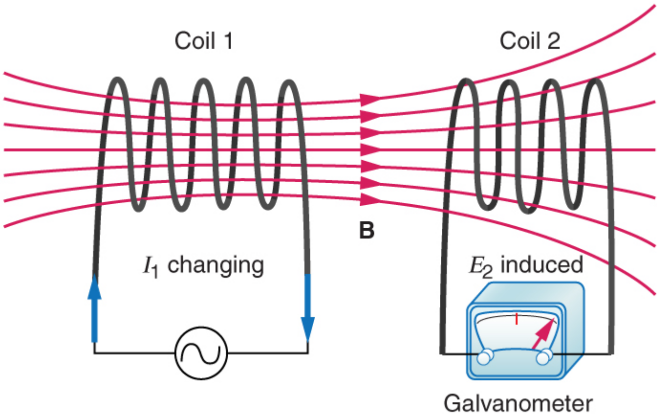
Inductance

- Calculate the inductance of an inductor.
- Calculate the energy stored in an inductor.
- Calculate the emf generated in an inductor.

Inductors

Induction is the process in which an emf is induced by changing magnetic flux. Many examples have been discussed so far, some more effective than others. Transformers, for example, are designed to be particularly effective at inducing a desired voltage and current with very little loss of energy to other forms. Is there a useful physical quantity related to how “effective” a given device is? The answer is yes, and that physical quantity is called **inductance**.

Mutual inductance is the effect of Faraday’s law of induction for one device upon another, such as the primary coil in transmitting energy to the secondary in a transformer. See [link], where simple coils induce emfs in one another.



These coils can induce emfs in one another like an inefficient transformer. Their mutual inductance M indicates the effectiveness of the coupling between them. Here a change in current in coil 1 is seen to induce an emf in coil 2. (Note that "E_2 induced" represents the induced emf in coil 2.)

In the many cases where the geometry of the devices is fixed, flux is changed by varying current. We therefore concentrate on the rate of change of current, $\Delta I/\Delta t$, as the cause of induction. A change in the current I_1 in one device, coil 1 in the figure, induces an emf_2 in the other. We express this in equation form as
Equation:

$$\text{emf}_2 = -M\frac{\Delta I_1}{\Delta t},$$

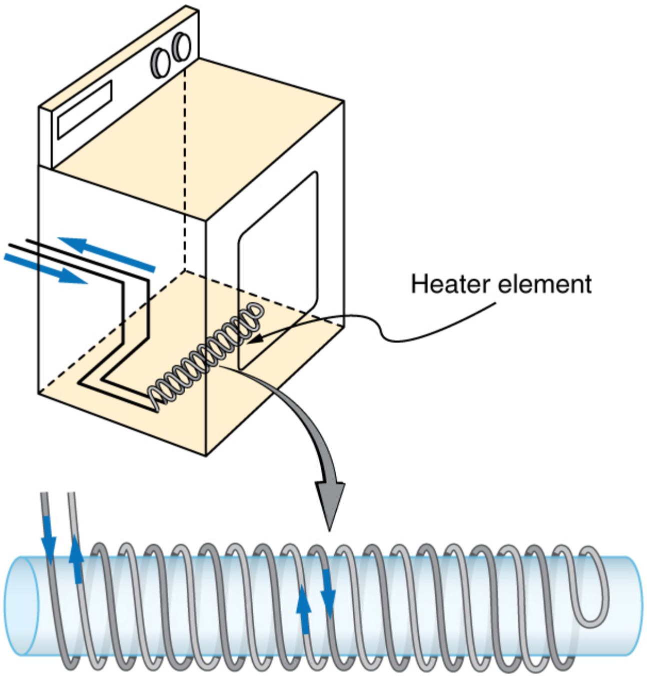
where M is defined to be the mutual inductance between the two devices. The minus sign is an expression of Lenz's law. The larger the mutual inductance M, the more effective the coupling. For example, the coils in [link] have a small M compared with the transformer coils in [link]. Units for M are $(\text{V}\cdot\text{s})/\text{A} = \Omega\cdot\text{s}$, which is named a **henry** (H), after Joseph Henry. That is, $1\text{ H} = 1\ \Omega\cdot\text{s}$.

Nature is symmetric here. If we change the current I_2 in coil 2, we induce an emf_1 in coil 1, which is given by
Equation:

$$\text{emf}_1 = -M\frac{\Delta I_2}{\Delta t},$$

where M is the same as for the reverse process. Transformers run backward with the same effectiveness, or mutual inductance M.

A large mutual inductance M may or may not be desirable. We want a transformer to have a large mutual inductance. But an appliance, such as an electric clothes dryer, can induce a dangerous emf on its case if the mutual inductance between its coils and the case is large. One way to reduce mutual inductance M is to counterwind coils to cancel the magnetic field produced. (See [link].)



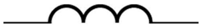
The heating coils of an electric clothes dryer can be counter-wound so that their magnetic fields cancel one another, greatly reducing the mutual inductance with the case of the dryer.

Self-inductance, the effect of Faraday’s law of induction of a device on itself, also exists. When, for example, current through a coil is increased, the magnetic field and flux also increase, inducing a counter emf, as required by Lenz's law. Conversely, if the current is decreased, an emf is induced that opposes the decrease. Most devices have a fixed geometry, and so the change in flux is due entirely to the change in current ΔI through the device. The induced emf is related to the physical geometry of the device and the rate of change of current. It is given by

Equation:

$$\text{emf} = -L\frac{\Delta I}{\Delta t},$$

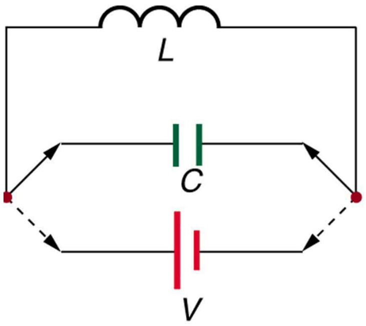
where L is the self-inductance of the device. A device that exhibits significant self-inductance is called an **inductor**, and given the symbol in [link].



The minus sign is an expression of Lenz's law, indicating that emf opposes the change in current. Units of self-inductance are henries (H) just as for mutual inductance. The larger the self-inductance L of a device, the greater its opposition to any change in current through it. For example, a large coil with many turns and an iron core has a large L and will not allow current to change quickly. To avoid this effect, a small L must be achieved, such as by counterwinding coils as in [link].

A 1 H inductor is a large inductor. To illustrate this, consider a device with $L = 1.0\text{ H}$ that has a 10 A current flowing through it. What happens if we try to shut off the current rapidly, perhaps in only 1.0 ms? An emf, given by $\text{emf} = -L(\Delta I/\Delta t)$, will oppose the change. Thus an emf will be induced given by $\text{emf} = -L(\Delta I/\Delta t) = (1.0\text{ H})[(10\text{ A})/(1.0\text{ ms})] = 10{,}000\text{ V}$. The positive sign means this large voltage is in the same direction as the current, opposing its decrease. Such large emfs can cause arcs, damaging switching equipment, and so it may be necessary to change current more slowly.

There are uses for such a large induced voltage. Camera flashes use a battery, two inductors that function as a transformer, and a switching system or oscillator to induce large voltages. (Remember that we need a changing magnetic field, brought about by a changing current, to induce a voltage in another coil.) The oscillator system will do this many times as the battery voltage is boosted to over one thousand volts. (You may hear the high pitched whine from the transformer as the capacitor is being charged.) A capacitor stores the high voltage for later use in powering the flash. (See [link].)



Through rapid switching of an inductor, 1.5 V batteries can be used to induce emfs of several thousand volts. This voltage can be used to store charge in a capacitor for later use, such as in a camera flash attachment.

It is possible to calculate L for an inductor given its geometry (size and shape) and knowing the magnetic field that it produces. This is difficult in most cases, because of the complexity of the field created. So in this text the inductance L is usually a given quantity. One exception is the solenoid, because it has a very uniform field inside, a nearly zero field outside, and a simple shape. It is instructive to derive an equation for its inductance. We start by noting that the induced emf is given by Faraday's law of induction as $\text{emf} = -N(\Delta\Phi/\Delta t)$ and, by the definition of self-inductance, as $\text{emf} = -L(\Delta I/\Delta t)$. Equating these yields

Equation:

$$\text{emf} = -N\frac{\Delta\Phi}{\Delta t} = -L\frac{\Delta I}{\Delta t}.$$

Solving for L gives

Equation:

$$L = N\frac{\Delta\Phi}{\Delta I}.$$

This equation for the self-inductance L of a device is always valid. It means that self-inductance L depends on how effective the current is in creating flux; the more effective, the greater $\Delta\Phi$/ ΔI is.

Let us use this last equation to find an expression for the inductance of a solenoid. Since the area A of a solenoid is fixed, the change in flux is $\Delta\Phi = \Delta(BA) = A\Delta B$. To find ΔB, we note that the magnetic field of a solenoid is given by $B = \mu_0\text{nI} = \mu_0\frac{\text{NI}}{\ell}$. (Here $n = N/\ell$, where N is the number of coils and ℓ is the solenoid's length.) Only the current changes, so that $\Delta\Phi = A\Delta B = \mu_0\text{NA}\frac{\Delta I}{\ell}$. Substituting $\Delta\Phi$ into $L = N\frac{\Delta\Phi}{\Delta I}$ gives

Equation:

$$L = N\frac{\Delta\Phi}{\Delta I} = N\frac{\mu_0\text{NA}\frac{\Delta I}{\ell}}{\Delta I}.$$

This simplifies to

Equation:

$$L = \frac{\mu_0 N^2 A}{\ell} \text{(solenoid)}.$$

This is the self-inductance of a solenoid of cross-sectional area A and length ℓ. Note that the inductance depends only on the physical characteristics of the solenoid, consistent with its definition.

Example:

Calculating the Self-inductance of a Moderate Size Solenoid

Calculate the self-inductance of a 10.0 cm long, 4.00 cm diameter solenoid that has 200 coils.

Strategy

This is a straightforward application of $L = \frac{\mu_0 N^2 A}{\ell}$, since all quantities in the equation except L are known.

Solution

Use the following expression for the self-inductance of a solenoid:

Equation:

$$L = \frac{\mu_0 N^2 A}{\ell}.$$

The cross-sectional area in this example is $A = \pi r^2 = (3.14...)(0.0200\text{ m})^2 = 1.26 \times 10^{-3}\text{ m}^2$, N is given to be 200, and the length ℓ is 0.100 m. We know the permeability of free space is $\mu_0 = 4\pi \times 10^{-7}\text{ T} \cdot \text{m/A}$. Substituting these into the expression for L gives

Equation:

$$\begin{aligned} L &= \frac{(4\pi \times 10^{-7}\text{ T}\cdot\text{m/A})(200)^2(1.26 \times 10^{-3}\text{ m}^2)}{0.100\text{ m}} \\ &= 0.632\text{ mH}. \end{aligned}$$

Discussion
This solenoid is moderate in size. Its inductance of nearly a millihenry is also considered moderate.

One common application of inductance is used in traffic lights that can tell when vehicles are waiting at the intersection. An electrical circuit with an inductor is placed in the road under the place a waiting car will stop over. The body of the car increases the inductance and the circuit changes sending a signal to the traffic lights to change colors. Similarly, metal detectors used for airport security employ the same technique. A coil or inductor in the metal detector frame acts as both a transmitter and a receiver. The pulsed signal in the transmitter coil induces a signal in the receiver. The self-inductance of the circuit is affected by any metal object in the path. Such detectors can be adjusted for sensitivity and also can indicate the approximate location of metal found on a person. (But they will not be able to detect any plastic explosive such as that found on the “underwear bomber.”) See [link].

The familiar security gate at an airport can not only detect metals but also indicate their approximate height above the floor.

(credit: Alexbuirds,
Wikimedia Commons)

Energy Stored in an Inductor

We know from Lenz's law that inductances oppose changes in current. There is an alternative way to look at this opposition that is based on energy. Energy is stored in a magnetic field. It takes time to build up energy, and it also takes time to deplete energy; hence, there is an opposition to rapid change. In an inductor, the magnetic field is directly proportional to current and to the inductance of the device. It can be shown that the **energy stored in an inductor** E_{ind} is given by

Equation:

$$E_{\text{ind}} = \frac{1}{2}LI^2.$$

This expression is similar to that for the energy stored in a capacitor.

Example:

Calculating the Energy Stored in the Field of a Solenoid

How much energy is stored in the 0.632 mH inductor of the preceding example when a 30.0 A current flows through it?

Strategy

The energy is given by the equation $E_{\text{ind}} = \frac{1}{2}LI^2$, and all quantities except E_{ind} are known.

Solution

Substituting the value for L found in the previous example and the given current into $E_{\text{ind}} = \frac{1}{2}LI^2$ gives

Equation:

$$\begin{aligned} E_{\text{ind}} &= \tfrac{1}{2}LI^2 \\ &= 0.5(0.632 \times 10^{-3}\text{ H})(30.0\text{ A})^2 = 0.284\text{ J}. \end{aligned}$$

Discussion
This amount of energy is certainly enough to cause a spark if the current is suddenly switched off. It cannot be built up instantaneously unless the power input is infinite.

Section Summary

- Inductance is the property of a device that tells how effectively it induces an emf in another device.
- Mutual inductance is the effect of two devices in inducing emfs in each other.
- A change in current $\Delta I_1/\Delta t$ in one induces an emf emf_2 in the second:
  Equation:
  $$\text{emf}_2 = -M\frac{\Delta I_1}{\Delta t},$$
  where M is defined to be the mutual inductance between the two devices, and the minus sign is due to Lenz's law.
- Symmetrically, a change in current $\Delta I_2/\Delta t$ through the second device induces an emf emf_1 in the first:
  Equation:
  $$\text{emf}_1 = -M\frac{\Delta I_2}{\Delta t},$$
  where M is the same mutual inductance as in the reverse process.
- Current changes in a device induce an emf in the device itself.
- Self-inductance is the effect of the device inducing emf in itself.
- The device is called an inductor, and the emf induced in it by a change in current through it is
  Equation:
  $$\text{emf} = -L\frac{\Delta I}{\Delta t},$$

where L is the self-inductance of the inductor, and $\Delta I/\Delta t$ is the rate of change of current through it. The minus sign indicates that emf opposes the change in current, as required by Lenz's law.

- The unit of self- and mutual inductance is the henry (H), where $1\ \text{H} = 1\ \Omega \cdot \text{s}$.
- The self-inductance L of an inductor is proportional to how much flux changes with current. For an N-turn inductor,

  Equation:

  $$L = N\frac{\Delta \Phi}{\Delta I}.$$

- The self-inductance of a solenoid is

  Equation:

  $$L = \frac{\mu_0 N^2 A}{\ell}\ \text{(solenoid)},$$

  where N is its number of turns in the solenoid, A is its cross-sectional area, ℓ is its length, and $\mu_0 = 4\pi \times 10^{-7}\ \text{T} \cdot \text{m/A}$ is the permeability of free space.
- The energy stored in an inductor E_{ind} is

  Equation:

  $$E_{\text{ind}} = \frac{1}{2}LI^2.$$

Conceptual Questions

Exercise:

Problem:

How would you place two identical flat coils in contact so that they had the greatest mutual inductance? The least?

Exercise:

Problem:

How would you shape a given length of wire to give it the greatest self-inductance? The least?

Exercise:

Problem:

Verify, as was concluded without proof in [link], that units of $\text{T} \cdot \text{m}^2 / A = \Omega \cdot \text{s} = \text{H}$.

Problems & Exercises

Exercise:

Problem:

Two coils are placed close together in a physics lab to demonstrate Faraday's law of induction. A current of 5.00 A in one is switched off in 1.00 ms, inducing a 9.00 V emf in the other. What is their mutual inductance?

Solution:

1.80 mH

Exercise:

Problem:

If two coils placed next to one another have a mutual inductance of 5.00 mH, what voltage is induced in one when the 2.00 A current in the other is switched off in 30.0 ms?

Exercise:

Problem:

The 4.00 A current through a 7.50 mH inductor is switched off in 8.33 ms. What is the emf induced opposing this?

Solution:

3.60 V

Exercise:

Problem:

A device is turned on and 3.00 A flows through it 0.100 ms later. What is the self-inductance of the device if an induced 150 V emf opposes this?

Exercise:

Problem:

Starting with $\text{emf}_2 = -M\frac{\Delta I_1}{\Delta t}$, show that the units of inductance are $(\text{V}\cdot\text{s})/\text{A} = \Omega\cdot\text{s}$.

Exercise:

Problem:

Camera flashes charge a capacitor to high voltage by switching the current through an inductor on and off rapidly. In what time must the 0.100 A current through a 2.00 mH inductor be switched on or off to induce a 500 V emf?

Exercise:

Problem:

A large research solenoid has a self-inductance of 25.0 H. (a) What induced emf opposes shutting it off when 100 A of current through it is switched off in 80.0 ms? (b) How much energy is stored in the inductor at full current? (c) At what rate in watts must energy be dissipated to switch the current off in 80.0 ms? (d) In view of the answer to the last part, is it surprising that shutting it down this quickly is difficult?

Solution:

(a) 31.3 kV

(b) 125 kJ

(c) 1.56 MW

(d) No, it is not surprising since this power is very high.

Exercise:

Problem:

(a) Calculate the self-inductance of a 50.0 cm long, 10.0 cm diameter solenoid having 1000 loops. (b) How much energy is stored in this inductor when 20.0 A of current flows through it? (c) How fast can it be turned off if the induced emf cannot exceed 3.00 V?

Exercise:

Problem:

A precision laboratory resistor is made of a coil of wire 1.50 cm in diameter and 4.00 cm long, and it has 500 turns. (a) What is its self-inductance? (b) What average emf is induced if the 12.0 A current through it is turned on in 5.00 ms (one-fourth of a cycle for 50 Hz AC)? (c) What is its inductance if it is shortened to half its length and counter-wound (two layers of 250 turns in opposite directions)?

Solution:

(a) 1.39 mH

(b) 3.33 V

(c) Zero

Exercise:

Problem:

The heating coils in a hair dryer are 0.800 cm in diameter, have a combined length of 1.00 m, and a total of 400 turns. (a) What is their total self-inductance assuming they act like a single solenoid? (b) How much energy is stored in them when 6.00 A flows? (c) What average emf opposes shutting them off if this is done in 5.00 ms (one-fourth of a cycle for 50 Hz AC)?

Exercise:

Problem:

When the 20.0 A current through an inductor is turned off in 1.50 ms, an 800 V emf is induced, opposing the change. What is the value of the self-inductance?

Solution:

60.0 mH

Exercise:

Problem:

How fast can the 150 A current through a 0.250 H inductor be shut off if the induced emf cannot exceed 75.0 V?

Exercise:

Problem: Integrated Concepts

A very large, superconducting solenoid such as one used in MRI scans, stores 1.00 MJ of energy in its magnetic field when 100 A flows. (a) Find its self-inductance. (b) If the coils “go normal,” they gain resistance and start to dissipate thermal energy. What temperature increase is produced if all the stored energy goes into heating the 1000 kg magnet, given its average specific heat is $200\ \text{J/kg}\cdot\text{ºC}$?

Solution:

(a) 200 H

(b) 5.00°C

Exercise:

Problem: **Unreasonable Results**

A 25.0 H inductor has 100 A of current turned off in 1.00 ms. (a) What voltage is induced to oppose this? (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

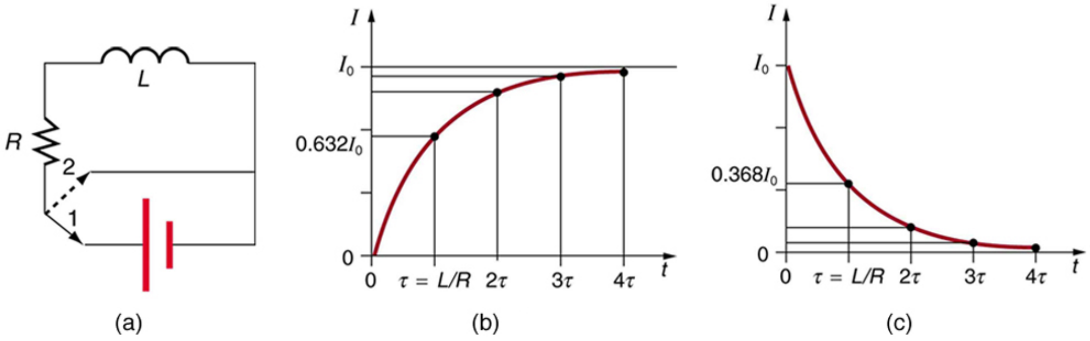
Glossary

inductance	a property of a device describing how efficient it is at inducing emf in another device
mutual inductance	how effective a pair of devices are at inducing emfs in each other
henry	the unit of inductance; $1\text{ H} = 1\ \Omega \cdot \text{s}$
self-inductance	how effective a device is at inducing emf in itself
inductor	a device that exhibits significant self-inductance
energy stored in an inductor	self-explanatory; calculated by $E_{\text{ind}} = \frac{1}{2}LI^2$

RL Circuits

- Calculate the current in an RL circuit after a specified number of characteristic time steps.
- Calculate the characteristic time of an RL circuit.
- Sketch the current in an RL circuit over time.

We know that the current through an inductor L cannot be turned on or off instantaneously. The change in current changes flux, inducing an emf opposing the change (Lenz’s law). How long does the opposition last? Current *will* flow and *can* be turned off, but how long does it take? [link] shows a switching circuit that can be used to examine current through an inductor as a function of time.



(a) An *RL* circuit with a switch to turn current on and off. When in position 1, the battery, resistor, and inductor are in series and a current is established. In position 2, the battery is removed and the current eventually stops because of energy loss in the resistor. (b) A graph of current growth versus time when the switch is moved to position 1. (c) A graph of current decay when the switch is moved to position 2.

When the switch is first moved to position 1 (at $t = 0$), the current is zero and it eventually rises to $I_0 = \mathrm{V/R}$, where R is the total resistance of the circuit. The opposition of the inductor L is greatest at the beginning, because the amount of change is greatest. The opposition it poses is in the form of an induced emf, which decreases to zero as the current approaches its final value. The opposing emf is proportional to the amount of change left. This is the hallmark of an exponential behavior, and it can be shown with calculus that

Equation:

$$I = I_0(1 - e^{-t/\tau}) \quad \text{(turning on)},$$

is the current in an *RL* circuit when switched on (Note the similarity to the exponential behavior of the voltage on a charging capacitor). The initial current is zero and approaches $I_0 = \text{V}/\text{R}$ with a **characteristic time constant** τ for an *RL* circuit, given by

Equation:

$$\tau = \frac{L}{R},$$

where τ has units of seconds, since 1 H=1 Ω·s. In the first period of time τ, the current rises from zero to $0.632I_0$, since $I = I_0(1 - e^{-1}) = I_0(1 - 0.368) = 0.632I_0$. The current will go 0.632 of the remainder in the next time τ. A well-known property of the exponential is that the final value is never exactly reached, but 0.632 of the remainder to that value is achieved in every characteristic time τ. In just a few multiples of the time τ, the final value is very nearly achieved, as the graph in [link](b) illustrates.

The characteristic time τ depends on only two factors, the inductance L and the resistance R. The greater the inductance L, the greater τ is, which makes sense since a large inductance is very effective in opposing change. The smaller the resistance R, the greater τ is. Again this makes sense, since a small resistance means a large final current and a greater change to get there. In both cases—large L and small R —more energy is stored in the inductor and more time is required to get it in and out.

When the switch in [link](a) is moved to position 2 and cuts the battery out of the circuit, the current drops because of energy dissipation by the resistor. But this is also not instantaneous, since the inductor opposes the decrease in current by inducing an emf in the same direction as the battery that drove the current. Furthermore, there is a certain amount of energy, $(1/2)LI_0^2$, stored in the inductor, and it is dissipated at a finite rate. As the current approaches zero, the rate of decrease slows, since the energy dissipation rate is I^2R. Once again the behavior is exponential, and I is found to be

Equation:

$$I = I_0e^{-t/\tau} \quad \text{(turning off)}.$$

(See [link](c).) In the first period of time $\tau = L/R$ after the switch is closed, the current falls to 0.368 of its initial value, since $I = I_0e^{-1} = 0.368I_0$. In each successive time τ, the current falls to 0.368 of the preceding value, and in a few multiples of τ, the current becomes very close to zero, as seen in the graph in [link](c).

Example:

Calculating Characteristic Time and Current in an *RL* Circuit

(a) What is the characteristic time constant for a 7.50 mH inductor in series with a 3.00 Ω resistor? (b) Find the current 5.00 ms after the switch is moved to position 2 to disconnect the battery, if it is initially 10.0 A.

Strategy for (a)

The time constant for an *RL* circuit is defined by $\tau = L/R$.

Solution for (a)

Entering known values into the expression for τ given in $\tau = L/R$ yields

Equation:

$$\tau = \frac{L}{R} = \frac{7.50 \text{ mH}}{3.00 \text{ }\Omega} = 2.50 \text{ ms}.$$

Discussion for (a)

This is a small but definitely finite time. The coil will be very close to its full current in about ten time constants, or about 25 ms.

Strategy for (b)

We can find the current by using $I = I_0e^{-t/\tau}$, or by considering the decline in steps. Since the time is twice the characteristic time, we consider the process in steps.

Solution for (b)
In the first 2.50 ms, the current declines to 0.368 of its initial value, which is

Equation:

$$\begin{aligned} I &= 0.368 I_0 = (0.368)(10.0\text{ A}) \\ &= 3.68\text{ A at } t = 2.50\text{ ms.} \end{aligned}$$

After another 2.50 ms, or a total of 5.00 ms, the current declines to 0.368 of the value just found. That is,

Equation:

$$\begin{aligned} I' &= 0.368 I = (0.368)(3.68\text{ A}) \\ &= 1.35\text{ A at } t = 5.00\text{ ms.} \end{aligned}$$

Discussion for (b)
After another 5.00 ms has passed, the current will be 0.183 A (see [link]); so, although it does die out, the current certainly does not go to zero instantaneously.

In summary, when the voltage applied to an inductor is changed, the current also changes, *but the change in current lags the change in voltage in an RL circuit*. In Reactance, Inductive and Capacitive, we explore how an *RL* circuit behaves when a sinusoidal AC voltage is applied.

Section Summary

- When a series connection of a resistor and an inductor—an *RL* circuit—is connected to a voltage source, the time variation of the current is

  Equation:

  $$I = I_0(1 - e^{-t/\tau}) \quad \text{(turning on).}$$

  where $I_0 = V/R$ is the final current.

- The characteristic time constant τ is $\tau = \frac{L}{R}$, where L is the inductance and R is the resistance.
- In the first time constant τ, the current rises from zero to $0.632I_0$, and 0.632 of the remainder in every subsequent time interval τ.
- When the inductor is shorted through a resistor, current decreases as **Equation:**

  $$I = I_0e^{-t/\tau} \quad (\text{turning off}).$$

  Here I_0 is the initial current.
- Current falls to $0.368I_0$ in the first time interval τ, and 0.368 of the remainder toward zero in each subsequent time τ.

Problem Exercises

Exercise:

Problem:

If you want a characteristic *RL* time constant of 1.00 s, and you have a $500\ \Omega$ resistor, what value of self-inductance is needed?

Solution:

500 H

Exercise:

Problem:

Your *RL* circuit has a characteristic time constant of 20.0 ns, and a resistance of $5.00\ \text{M}\Omega$. (a) What is the inductance of the circuit? (b) What resistance would give you a 1.00 ns time constant, perhaps needed for quick response in an oscilloscope?

Exercise:

Problem:

A large superconducting magnet, used for magnetic resonance imaging, has a 50.0 H inductance. If you want current through it to be adjustable with a 1.00 s characteristic time constant, what is the minimum resistance of system?

Solution:

$50.0\ \Omega$

Exercise:

Problem:

Verify that after a time of 10.0 ms, the current for the situation considered in [link] will be 0.183 A as stated.

Exercise:

Problem:

Suppose you have a supply of inductors ranging from 1.00 nH to 10.0 H, and resistors ranging from $0.100\ \Omega$ to $1.00\ \text{M}\Omega$. What is the range of characteristic *RL* time constants you can produce by connecting a single resistor to a single inductor?

Solution:

1.00×10^{-18} s to 0.100 s

Exercise:

Problem:

(a) What is the characteristic time constant of a 25.0 mH inductor that has a resistance of $4.00\ \Omega$? (b) If it is connected to a 12.0 V battery, what is the current after 12.5 ms?

Exercise:

Problem:

What percentage of the final current I_0 flows through an inductor L in series with a resistor R, three time constants after the circuit is completed?

Solution:

95.0%

Exercise:

Problem:

The 5.00 A current through a 1.50 H inductor is dissipated by a $2.00\ \Omega$ resistor in a circuit like that in [link] with the switch in position 2. (a) What is the initial energy in the inductor? (b) How long will it take the current to decline to 5.00% of its initial value? (c) Calculate the average power dissipated, and compare it with the initial power dissipated by the resistor.

Exercise:

Problem:

(a) Use the exact exponential treatment to find how much time is required to bring the current through an 80.0 mH inductor in series with a $15.0\ \Omega$ resistor to 99.0% of its final value, starting from zero. (b) Compare your answer to the approximate treatment using integral numbers of τ. (c) Discuss how significant the difference is.

Solution:

(a) 24.6 ms

(b) 26.7 ms

(c) 9% difference, which is greater than the inherent uncertainty in the given parameters.

Exercise:

Problem:

(a) Using the exact exponential treatment, find the time required for the current through a 2.00 H inductor in series with a $0.500\ \Omega$ resistor to be reduced to 0.100% of its original value. (b) Compare your answer to the approximate treatment using integral numbers of τ. (c) Discuss how significant the difference is.

Glossary

characteristic time constant

denoted by τ, of a particular series *RL* circuit is calculated by $\tau = \frac{L}{R}$, where L is the inductance and R is the resistance

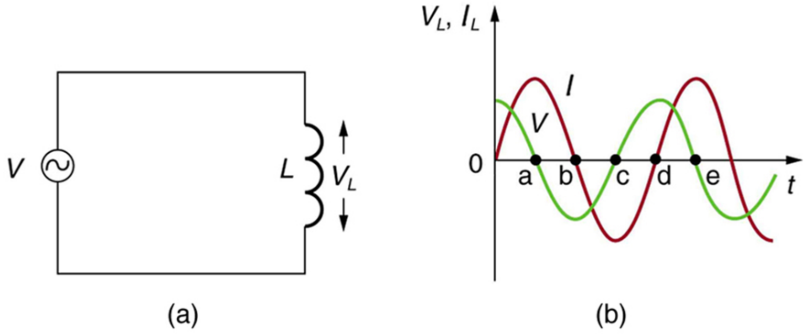
Reactance, Inductive and Capacitive

- Sketch voltage and current versus time in simple inductive, capacitive, and resistive circuits.
- Calculate inductive and capacitive reactance.
- Calculate current and/or voltage in simple inductive, capacitive, and resistive circuits.

Many circuits also contain capacitors and inductors, in addition to resistors and an AC voltage source. We have seen how capacitors and inductors respond to DC voltage when it is switched on and off. We will now explore how inductors and capacitors react to sinusoidal AC voltage.

Inductors and Inductive Reactance

Suppose an inductor is connected directly to an AC voltage source, as shown in [link]. It is reasonable to assume negligible resistance, since in practice we can make the resistance of an inductor so small that it has a negligible effect on the circuit. Also shown is a graph of voltage and current as functions of time.



(a) An AC voltage source in series with an inductor having negligible resistance. (b) Graph of current and voltage across the inductor as functions of time.

The graph in [link](b) starts with voltage at a maximum. Note that the current starts at zero and rises to its peak *after* the voltage that drives it, just as was the case when DC voltage was switched on in the preceding section. When the voltage becomes negative at point a, the current begins to decrease; it becomes zero at point b, where voltage is its most negative. The current then becomes negative, again following the voltage. The voltage becomes positive at point c and begins to make the current less negative. At point d, the current goes through zero just as the voltage reaches its positive peak to start another cycle. This behavior is summarized as follows:

Note:
AC Voltage in an Inductor
When a sinusoidal voltage is applied to an inductor, the voltage leads the current by one-fourth of a cycle, or by a 90º phase angle.

Current lags behind voltage, since inductors oppose change in current. Changing current induces a back emf $V = -L(\Delta I/\Delta t)$. This is considered to be an effective resistance of the inductor to AC. The rms current I through an inductor L is given by a version of Ohm's law:
Equation:

$$I = \frac{V}{X_L},$$

where V is the rms voltage across the inductor and X_L is defined to be
Equation:

$$X_L = 2\pi fL,$$

with f the frequency of the AC voltage source in hertz (An analysis of the circuit using Kirchhoff's loop rule and calculus actually produces this expression). X_L is called the **inductive reactance**, because the inductor reacts to impede the current. X_L has units of ohms ($1\ \mathrm{H} = 1\ \Omega \cdot \mathrm{s}$, so that frequency times inductance has units of $(\mathrm{cycles/s})(\Omega \cdot \mathrm{s}) = \Omega$), consistent with its role as an effective resistance. It makes sense that X_L is proportional to L, since the greater the induction the greater its resistance to change. It is also reasonable that X_L is proportional to frequency f, since greater frequency means greater change in current. That is, $\Delta I/\Delta t$ is large for large frequencies (large f, small Δt). The greater the change, the greater the opposition of an inductor.

Example:
Calculating Inductive Reactance and then Current
(a) Calculate the inductive reactance of a 3.00 mH inductor when 60.0 Hz and 10.0 kHz AC voltages are applied. (b) What is the rms current at each frequency if the applied rms voltage is 120 V?
Strategy
The inductive reactance is found directly from the expression $X_L = 2\pi\mathrm{fL}$. Once X_L has been found at each frequency, Ohm’s law as stated in the Equation $I = V/X_L$ can be used to find the current at each frequency.
Solution for (a)
Entering the frequency and inductance into Equation $X_L = 2\pi\mathrm{fL}$ gives
Equation:

$$X_L = 2\pi\mathrm{fL} = 6.28(60.0/\mathrm{s})(3.00\ \mathrm{mH}) = 1.13\ \Omega \text{ at } 60\ \mathrm{Hz}.$$

Similarly, at 10 kHz,
Equation:

$$X_L = 2\pi\mathrm{fL} = 6.28(1.00 \times 10^4/\mathrm{s})(3.00\ \mathrm{mH}) = 188\ \Omega \text{ at } 10\ \mathrm{kHz}.$$

Solution for (b)
The rms current is now found using the version of Ohm’s law in Equation $I = V/X_L$, given the applied rms voltage is 120 V. For the first frequency, this yields
Equation:

$$I = \frac{V}{X_L} = \frac{120\text{ V}}{1.13\ \Omega} = 106\text{ A at }60\text{ Hz.}$$

Similarly, at 10 kHz,

Equation:

$$I = \frac{V}{X_L} = \frac{120\text{ V}}{188\ \Omega} = 0.637\text{ A at }10\text{ kHz.}$$

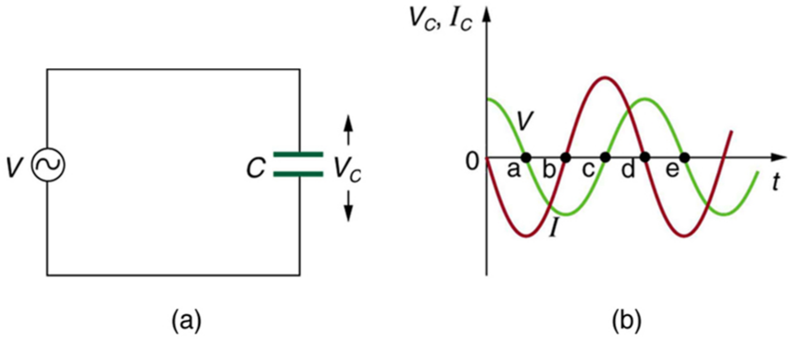
Discussion

The inductor reacts very differently at the two different frequencies. At the higher frequency, its reactance is large and the current is small, consistent with how an inductor impedes rapid change. Thus high frequencies are impeded the most. Inductors can be used to filter out high frequencies; for example, a large inductor can be put in series with a sound reproduction system or in series with your home computer to reduce high-frequency sound output from your speakers or high-frequency power spikes into your computer.

Note that although the resistance in the circuit considered is negligible, the AC current is not extremely large because inductive reactance impedes its flow. With AC, there is no time for the current to become extremely large.

Capacitors and Capacitive Reactance

Consider the capacitor connected directly to an AC voltage source as shown in [link]. The resistance of a circuit like this can be made so small that it has a negligible effect compared with the capacitor, and so we can assume negligible resistance. Voltage across the capacitor and current are graphed as functions of time in the figure.



(a) An AC voltage source in series with a capacitor *C* having negligible resistance. (b) Graph of current and voltage across the capacitor as functions of time.

The graph in [link] starts with voltage across the capacitor at a maximum. The current is zero at this point, because the capacitor is fully charged and halts the flow. Then voltage drops and the current becomes negative as the capacitor discharges. At point a, the capacitor has fully discharged ($Q = 0$ on it) and the voltage across it is zero. The current remains negative between points a and b, causing the voltage on the capacitor to reverse. This is complete at point b, where the current is zero and the voltage has its most negative value. The current becomes positive after point b, neutralizing the charge on the capacitor and bringing the voltage to zero at point c, which allows the current to reach its maximum. Between points c and d, the current drops to zero as the voltage rises to its peak, and the process starts to repeat. Throughout the cycle, the voltage follows what the current is doing by one-fourth of a cycle:

Note:

AC Voltage in a Capacitor

When a sinusoidal voltage is applied to a capacitor, the voltage follows the current by one-fourth of a cycle, or by a 90° phase angle.

The capacitor is affecting the current, having the ability to stop it altogether when fully charged. Since an AC voltage is applied, there is an rms current, but it is limited by the capacitor. This is considered to be an effective resistance of the capacitor to AC, and so the rms current I in the circuit containing only a capacitor C is given by another version of Ohm's law to be

Equation:

$$I = \frac{V}{X_C},$$

where V is the rms voltage and X_C is defined (As with X_L, this expression for X_C results from an analysis of the circuit using Kirchhoff's rules and calculus) to be

Equation:

$$X_C = \frac{1}{2\pi fC},$$

where X_C is called the **capacitive reactance**, because the capacitor reacts to impede the current. X_C has units of ohms (verification left as an exercise for the reader). X_C is inversely proportional to the capacitance C; the larger the capacitor, the greater the charge it can store and the greater the current that can flow. It is also inversely proportional to the frequency f; the greater the frequency, the less time there is to fully charge the capacitor, and so it impedes current less.

Example:

Calculating Capacitive Reactance and then Current

(a) Calculate the capacitive reactance of a 5.00 mF capacitor when 60.0 Hz and 10.0 kHz AC voltages are applied. (b) What is the rms current if the applied rms voltage is 120 V?

Strategy

The capacitive reactance is found directly from the expression in $X_C = \frac{1}{2\pi fC}$. Once X_C has been found at each frequency, Ohm's law stated as $I = V/X_C$ can be used to find the current at each frequency.

Solution for (a)

Entering the frequency and capacitance into $X_C = \frac{1}{2\pi fC}$ gives

Equation:

$$\begin{aligned} X_C &= \frac{1}{2\pi fC} \\ &= \frac{1}{6.28(60.0/\text{s})(5.00\ \mu\text{F})} = 531\ \Omega \text{ at } 60\text{ Hz.} \end{aligned}$$

Similarly, at 10 kHz,

Equation:

$$\begin{aligned} X_C &= \frac{1}{2\pi fC} = \frac{1}{6.28(1.00\times 10^4/\text{s})(5.00\ \mu\text{F})} \\ &= 3.18\ \Omega \text{ at } 10\text{ kHz} \end{aligned}.$$

Solution for (b)

The rms current is now found using the version of Ohm's law in $I = V/X_C$, given the applied rms voltage is 120 V. For the first frequency, this yields

Equation:

$$I = \frac{V}{X_C} = \frac{120\text{ V}}{531\ \Omega} = 0.226\text{ A at } 60\text{ Hz.}$$

Similarly, at 10 kHz,

Equation:

$$I = \frac{V}{X_C} = \frac{120\text{ V}}{3.18\ \Omega} = 37.7\text{ A at } 10\text{ kHz.}$$

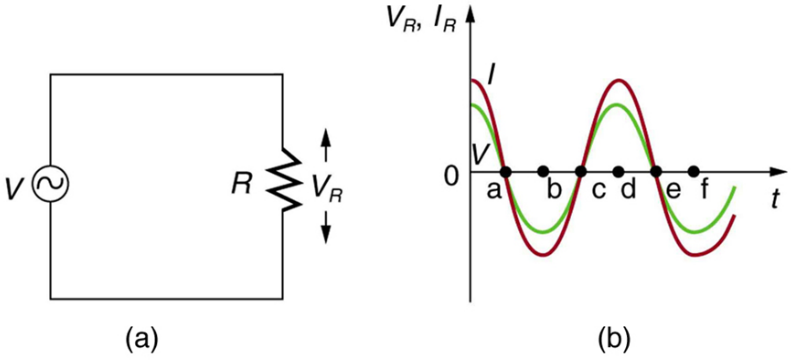
Discussion

The capacitor reacts very differently at the two different frequencies, and in exactly the opposite way an inductor reacts. At the higher frequency, its reactance is small and the current is large. Capacitors favor change, whereas inductors oppose change. Capacitors impede low frequencies the most, since low frequency allows them time to become charged and stop the current. Capacitors can be used to filter out low frequencies. For example, a capacitor in series with a sound reproduction system rids it of the 60 Hz hum.

Although a capacitor is basically an open circuit, there is an rms current in a circuit with an AC voltage applied to a capacitor. This is because the voltage is continually reversing, charging and discharging the capacitor. If the frequency goes to zero (DC), X_C tends to infinity, and the current is zero once the capacitor is charged. At very high frequencies, the capacitor's reactance tends to zero—it has a negligible reactance and does not impede the current (it acts like a simple wire). *Capacitors have the opposite effect on AC circuits that inductors have*.

Resistors in an AC Circuit

Just as a reminder, consider [link], which shows an AC voltage applied to a resistor and a graph of voltage and current versus time. The voltage and current are exactly *in phase* in a resistor. There is no frequency dependence to the behavior of plain resistance in a circuit:



(a) An AC voltage source in series with a resistor. (b) Graph of current and voltage across the resistor as functions of time, showing them to be exactly in phase.

Note:
AC Voltage in a Resistor
When a sinusoidal voltage is applied to a resistor, the voltage is exactly in phase with the current—they have a 0º phase angle.

Section Summary

- For inductors in AC circuits, we find that when a sinusoidal voltage is applied to an inductor, the voltage leads the current by one-fourth of a cycle, or by a 90º phase angle.
- The opposition of an inductor to a change in current is expressed as a type of AC resistance.
- Ohm’s law for an inductor is
  Equation:

  $$I=\frac{V}{X_L},$$

  where V is the rms voltage across the inductor.
- X_L is defined to be the inductive reactance, given by
  Equation:

  $$X_L=2\pi fL,$$

  with f the frequency of the AC voltage source in hertz.
- Inductive reactance X_L has units of ohms and is greatest at high frequencies.
- For capacitors, we find that when a sinusoidal voltage is applied to a capacitor, the voltage follows the current by one-fourth of a cycle, or

by a 90^{o} phase angle.

- Since a capacitor can stop current when fully charged, it limits current and offers another form of AC resistance; Ohm's law for a capacitor is
  Equation:

  $$I = \frac{V}{X_C},$$

  where V is the rms voltage across the capacitor.
- X_C is defined to be the capacitive reactance, given by
  Equation:

  $$X_C = \frac{1}{2\pi fC}.$$
- X_C has units of ohms and is greatest at low frequencies.

Conceptual Questions

Exercise:

Problem:

Presbycusis is a hearing loss due to age that progressively affects higher frequencies. A hearing aid amplifier is designed to amplify all frequencies equally. To adjust its output for presbycusis, would you put a capacitor in series or parallel with the hearing aid's speaker? Explain.

Exercise:

Problem:

Would you use a large inductance or a large capacitance in series with a system to filter out low frequencies, such as the 100 Hz hum in a sound system? Explain.

Exercise:

Problem:

High-frequency noise in AC power can damage computers. Does the plug-in unit designed to prevent this damage use a large inductance or a large capacitance (in series with the computer) to filter out such high frequencies? Explain.

Exercise:

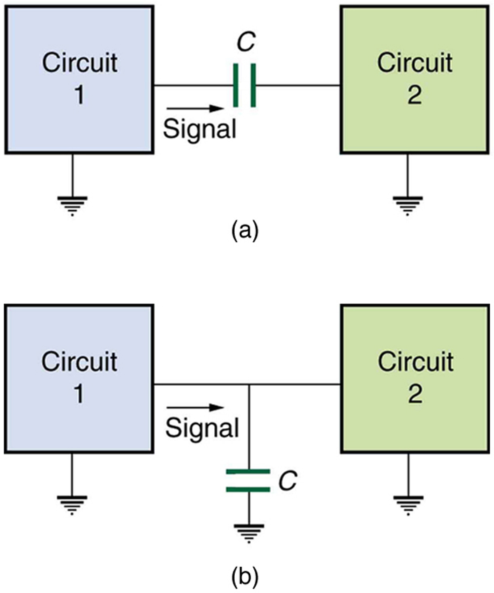
Problem:

Does inductance depend on current, frequency, or both? What about inductive reactance?

Exercise:

Problem:

Explain why the capacitor in [link](a) acts as a low-frequency filter between the two circuits, whereas that in [link](b) acts as a high-frequency filter.



Capacitors and inductors.

Capacitor with high frequency and low frequency.

Exercise:

Problem:

If the capacitors in [link] are replaced by inductors, which acts as a low-frequency filter and which as a high-frequency filter?

Problems & Exercises

Exercise:

Problem:

At what frequency will a 30.0 mH inductor have a reactance of $100\ \Omega$?

Solution:

531 Hz

Exercise:

Problem:

What value of inductance should be used if a $20.0\ \text{k}\Omega$ reactance is needed at a frequency of 500 Hz?

Exercise:

Problem:

What capacitance should be used to produce a $2.00\ \text{M}\Omega$ reactance at 60.0 Hz?

Solution:

1.33 nF

Exercise:

Problem:

At what frequency will an 80.0 mF capacitor have a reactance of $0.250\ \Omega$?

Exercise:

Problem:

(a) Find the current through a 0.500 H inductor connected to a 60.0 Hz, 480 V AC source. (b) What would the current be at 100 kHz?

Solution:

(a) 2.55 A

(b) 1.53 mA

Exercise:

Problem:

(a) What current flows when a 60.0 Hz, 480 V AC source is connected to a $0.250\ \mu F$ capacitor? (b) What would the current be at 25.0 kHz?

Exercise:

Problem:

A 20.0 kHz, 16.0 V source connected to an inductor produces a 2.00 A current. What is the inductance?

Solution:

63.7 µH

Exercise:

Problem:

A 20.0 Hz, 16.0 V source produces a 2.00 mA current when connected to a capacitor. What is the capacitance?

Exercise:

Problem:

(a) An inductor designed to filter high-frequency noise from power supplied to a personal computer is placed in series with the computer. What minimum inductance should it have to produce a $2.00\ \text{k}\Omega$ reactance for 15.0 kHz noise? (b) What is its reactance at 60.0 Hz?

Solution:

(a) 21.2 mH

(b) 8.00 Ω

Exercise:

Problem:

The capacitor in [link](a) is designed to filter low-frequency signals, impeding their transmission between circuits. (a) What capacitance is needed to produce a $100\ \text{k}\Omega$ reactance at a frequency of 120 Hz? (b) What would its reactance be at 1.00 MHz? (c) Discuss the implications of your answers to (a) and (b).

Exercise:

Problem:

The capacitor in [link](b) will filter high-frequency signals by shorting them to earth/ground. (a) What capacitance is needed to produce a reactance of $10.0\ \text{m}\Omega$ for a 5.00 kHz signal? (b) What would its reactance be at 3.00 Hz? (c) Discuss the implications of your answers to (a) and (b).

Solution:

(a) 3.18 mF

(b) 16.7 Ω

Exercise:

Problem: Unreasonable Results

In a recording of voltages due to brain activity (an EEG), a 10.0 mV signal with a 0.500 Hz frequency is applied to a capacitor, producing a current of 100 mA. Resistance is negligible. (a) What is the capacitance? (b) What is unreasonable about this result? (c) Which assumption or premise is responsible?

Exercise:

Problem: Construct Your Own Problem

Consider the use of an inductor in series with a computer operating on 60 Hz electricity. Construct a problem in which you calculate the relative reduction in voltage of incoming high frequency noise compared to 60 Hz voltage. Among the things to consider are the acceptable series reactance of the inductor for 60 Hz power and the likely frequencies of noise coming through the power lines.

Glossary

inductive reactance
: the opposition of an inductor to a change in current; calculated by $X_L = 2\pi \mathrm{fL}$

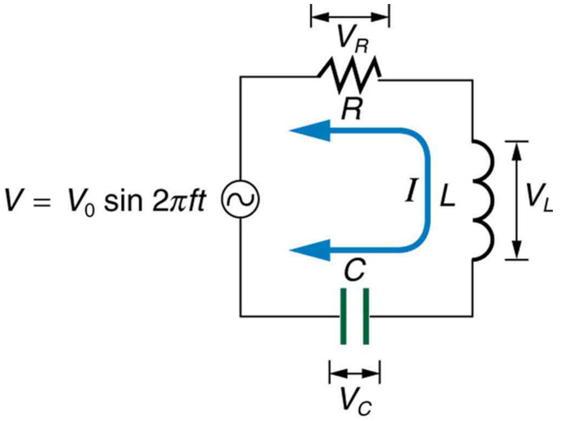
capacitive reactance
: the opposition of a capacitor to a change in current; calculated by $X_C = \frac{1}{2\pi \mathrm{fC}}$

RLC Series AC Circuits

- Calculate the impedance, phase angle, resonant frequency, power, power factor, voltage, and/or current in a RLC series circuit.
- Draw the circuit diagram for an RLC series circuit.
- Explain the significance of the resonant frequency.

Impedance

When alone in an AC circuit, inductors, capacitors, and resistors all impede current. How do they behave when all three occur together? Interestingly, their individual resistances in ohms do not simply add. Because inductors and capacitors behave in opposite ways, they partially to totally cancel each other's effect. [link] shows an *RLC* series circuit with an AC voltage source, the behavior of which is the subject of this section. The crux of the analysis of an *RLC* circuit is the frequency dependence of X_L and X_C, and the effect they have on the phase of voltage versus current (established in the preceding section). These give rise to the frequency dependence of the circuit, with important "resonance" features that are the basis of many applications, such as radio tuners.



An *RLC* series circuit with an AC voltage source.

The combined effect of resistance R, inductive reactance X_L, and capacitive reactance X_C is defined to be **impedance**, an AC analogue to resistance in a DC circuit. Current, voltage, and impedance in an *RLC* circuit are related by an AC version of Ohm’s law:

Equation:

$$I_0 = \frac{V_0}{Z} \text{ or } I_{\text{rms}} = \frac{V_{\text{rms}}}{Z}.$$

Here I_0 is the peak current, V_0 the peak source voltage, and Z is the impedance of the circuit. The units of impedance are ohms, and its effect on the circuit is as you might expect: the greater the impedance, the smaller the current. To get an expression for Z in terms of R, X_L, and X_C, we will now examine how the voltages across the various components are related to the source voltage. Those voltages are labeled V_R, V_L, and V_C in [link].

Conservation of charge requires current to be the same in each part of the circuit at all times, so that we can say the currents in R, L, and C are equal and in phase. But we know from the preceding section that the voltage across the inductor V_L leads the current by one-fourth of a cycle, the voltage across the capacitor V_C follows the current by one-fourth of a cycle, and the voltage across the resistor V_R is exactly in phase with the current. [link] shows these relationships in one graph, as well as showing the total voltage around the circuit $V = V_R + V_L + V_C$, where all four voltages are the instantaneous values. According to Kirchhoff’s loop rule, the total voltage around the circuit V is also the voltage of the source.

You can see from [link] that while V_R is in phase with the current, V_L leads by 90º, and V_C follows by 90º. Thus V_L and V_C are 180º out of phase (crest to trough) and tend to cancel, although not completely unless they have the same magnitude. Since the peak voltages are not aligned (not in phase), the peak voltage V_0 of the source does *not* equal the sum of the peak voltages across R, L, and C. The actual relationship is

Equation:

$$V_0 = \sqrt{{V_{0R}}^2 + (V_{0L} - V_{0C})^2},$$

where V_{0R}, V_{0L}, and V_{0C} are the peak voltages across R, L, and C, respectively. Now, using Ohm's law and definitions from Reactance, Inductive and Capacitive, we substitute $V_0 = I_0 Z$ into the above, as well as $V_{0R} = I_0 R$, $V_{0L} = I_0 X_L$, and $V_{0C} = I_0 X_C$, yielding

Equation:

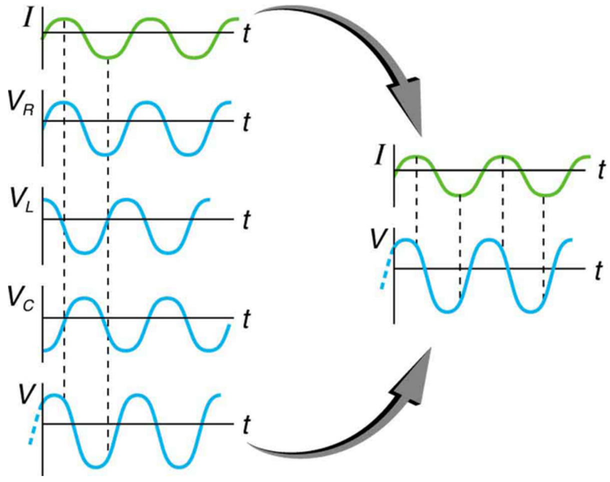
$$I_0 Z = \sqrt{{I_0}^2 R^2 + (I_0 X_L - I_0 X_C)^2} = I_0 \sqrt{R^2 + (X_L - X_C)^2}.$$

I_0 cancels to yield an expression for Z:

Equation:

$$Z = \sqrt{R^2 + (X_L - X_C)^2},$$

which is the impedance of an *RLC* series AC circuit. For circuits without a resistor, take $R = 0$; for those without an inductor, take $X_L = 0$; and for those without a capacitor, take $X_C = 0$.



This graph shows the relationships of the voltages in an *RLC* circuit to the current.

The voltages across the circuit elements add to equal the voltage of the source, which is seen to be out of phase with the current.

Example:

Calculating Impedance and Current

An *RLC* series circuit has a 40.0 Ω resistor, a 3.00 mH inductor, and a 5.00 μF capacitor. (a) Find the circuit’s impedance at 60.0 Hz and 10.0 kHz, noting that these frequencies and the values for L and C are the same as in [link] and [link]. (b) If the voltage source has $V_{\text{rms}} = 120\text{ V}$, what is I_{rms} at each frequency?

Strategy

For each frequency, we use $Z = \sqrt{R^2 + (X_L - X_C)^2}$ to find the impedance and then Ohm’s law to find current. We can take advantage of the results of the previous two examples rather than calculate the reactances again.

Solution for (a)

At 60.0 Hz, the values of the reactances were found in [link] to be $X_L = 1.13\ \Omega$ and in [link] to be $X_C = 531\ \Omega$. Entering these and the given 40.0 Ω for resistance into $Z = \sqrt{R^2 + (X_L - X_C)^2}$ yields

Equation:

$$\begin{aligned} Z &= \sqrt{R^2 + (X_L - X_C)^2} \\ &= \sqrt{(40.0\ \Omega)^2 + (1.13\ \Omega - 531\ \Omega)^2} \\ &= 531\ \Omega \text{ at } 60.0\text{ Hz}. \end{aligned}$$

Similarly, at 10.0 kHz, $X_L = 188\ \Omega$ and $X_C = 3.18\ \Omega$, so that

Equation:

$$\begin{aligned} Z &= \sqrt{(40.0\ \Omega)^2 + (188\ \Omega - 3.18\ \Omega)^2} \\ &= 190\ \Omega \text{ at } 10.0\text{ kHz.} \end{aligned}$$

Discussion for (a)

In both cases, the result is nearly the same as the largest value, and the impedance is definitely not the sum of the individual values. It is clear that X_L dominates at high frequency and X_C dominates at low frequency.

Solution for (b)

The current I_{rms} can be found using the AC version of Ohm's law in Equation $I_{\text{rms}} = V_{\text{rms}}/Z$:

$I_{\text{rms}} = \frac{V_{\text{rms}}}{Z} = \frac{120\text{ V}}{531\ \Omega} = 0.226\text{ A}$ at 60.0 Hz

Finally, at 10.0 kHz, we find

$I_{\text{rms}} = \frac{V_{\text{rms}}}{Z} = \frac{120\text{ V}}{190\ \Omega} = 0.633\text{ A}$ at 10.0 kHz

Discussion for (a)

The current at 60.0 Hz is the same (to three digits) as found for the capacitor alone in [link]. The capacitor dominates at low frequency. The current at 10.0 kHz is only slightly different from that found for the inductor alone in [link]. The inductor dominates at high frequency.

Resonance in *RLC* Series AC Circuits

How does an *RLC* circuit behave as a function of the frequency of the driving voltage source? Combining Ohm's law, $I_{\text{rms}} = V_{\text{rms}}/Z$, and the expression for impedance Z from $Z = \sqrt{R^2 + (X_L - X_C)^2}$ gives

Equation:

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{\sqrt{R^2 + (X_L - X_C)^2}}.$$

The reactances vary with frequency, with X_L large at high frequencies and X_C large at low frequencies, as we have seen in three previous examples. At some intermediate frequency f_0, the reactances will be equal and cancel, giving $Z = R$—this is a minimum value for impedance, and a maximum value for I_{rms} results. We can get an expression for f_0 by taking

Equation:

$$X_L = X_C.$$

Substituting the definitions of X_L and X_C,

Equation:

$$2\pi f_0 L = \frac{1}{2\pi f_0 C}.$$

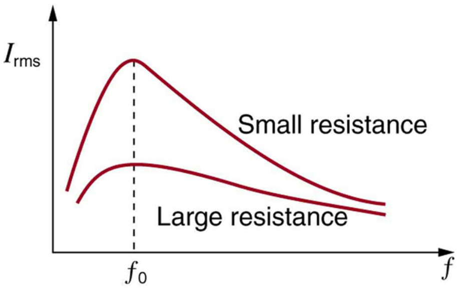
Solving this expression for f_0 yields

Equation:

$$f_0 = \frac{1}{2\pi\sqrt{\text{LC}}},$$

where f_0 is the **resonant frequency** of an *RLC* series circuit. This is also the *natural frequency* at which the circuit would oscillate if not driven by the voltage source. At f_0, the effects of the inductor and capacitor cancel, so that $Z = R$, and I_{rms} is a maximum.

Resonance in AC circuits is analogous to mechanical resonance, where resonance is defined to be a forced oscillation—in this case, forced by the voltage source—at the natural frequency of the system. The receiver in a radio is an *RLC* circuit that oscillates best at its f_0. A variable capacitor is often used to adjust f_0 to receive a desired frequency and to reject others. [link] is a graph of current as a function of frequency, illustrating a resonant peak in I_{rms} at f_0. The two curves are for two different circuits, which differ only in the amount of resistance in them. The peak is lower and broader for the higher-resistance circuit. Thus the higher-resistance circuit does not resonate as strongly and would not be as selective in a radio receiver, for example.



A graph of current versus frequency for two *RLC* series circuits differing only in the amount of resistance. Both have a resonance at f_0, but that for the higher resistance is lower and broader. The driving AC voltage source has a fixed amplitude V_0.

Example:

Calculating Resonant Frequency and Current

For the same *RLC* series circuit having a $40.0\ \Omega$ resistor, a 3.00 mH inductor, and a $5.00\ \mu\text{F}$ capacitor: (a) Find the resonant frequency. (b) Calculate I_{rms} at resonance if V_{rms} is 120 V.

Strategy

The resonant frequency is found by using the expression in $f_0 = \frac{1}{2\pi\sqrt{\text{LC}}}$.

The current at that frequency is the same as if the resistor alone were in the circuit.

Solution for (a)

Entering the given values for L and C into the expression given for f_0 in $f_0 = \frac{1}{2\pi\sqrt{\text{LC}}}$ yields

Equation:

$$\begin{aligned} f_0 &= \frac{1}{2\pi\sqrt{LC}} \\ &= \frac{1}{2\pi\sqrt{(3.00\times10^{-3}\text{ H})(5.00\times10^{-6}\text{ F})}} = 1.30\text{ kHz}. \end{aligned}$$

Discussion for (a)
We see that the resonant frequency is between 60.0 Hz and 10.0 kHz, the two frequencies chosen in earlier examples. This was to be expected, since the capacitor dominated at the low frequency and the inductor dominated at the high frequency. Their effects are the same at this intermediate frequency.

Solution for (b)
The current is given by Ohm’s law. At resonance, the two reactances are equal and cancel, so that the impedance equals the resistance alone. Thus,

Equation:

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{Z} = \frac{120\text{ V}}{40.0\ \Omega} = 3.00\text{ A}.$$

Discussion for (b)
At resonance, the current is greater than at the higher and lower frequencies considered for the same circuit in the preceding example.

Power in *RLC* Series AC Circuits

If current varies with frequency in an *RLC* circuit, then the power delivered to it also varies with frequency. But the average power is not simply current times voltage, as it is in purely resistive circuits. As was seen in [link], voltage and current are out of phase in an *RLC* circuit. There is a **phase angle** ϕ between the source voltage V and the current I, which can be found from

Equation:

$$\cos\,\phi = \frac{R}{Z}.$$

For example, at the resonant frequency or in a purely resistive circuit $Z = R$, so that $\cos\,\phi = 1$. This implies that $\phi = 0^{\text{o}}$ and that voltage and current are in phase, as expected for resistors. At other frequencies, average power is less than at resonance. This is both because voltage and current are out of phase and because I_{rms} is lower. The fact that source voltage and current are out of phase affects the power delivered to the circuit. It can be shown that the *average power* is

Equation:

$$P_{\text{ave}} = I_{\text{rms}} V_{\text{rms}} \cos\,\phi,$$

Thus $\cos\,\phi$ is called the **power factor**, which can range from 0 to 1. Power factors near 1 are desirable when designing an efficient motor, for example. At the resonant frequency, $\cos\,\phi = 1$.

Example:

Calculating the Power Factor and Power

For the same *RLC* series circuit having a 40.0 Ω resistor, a 3.00 mH inductor, a 5.00 μF capacitor, and a voltage source with a V_{rms} of 120 V: (a) Calculate the power factor and phase angle for $f = 60.0\text{Hz}$. (b) What is the average power at 50.0 Hz? (c) Find the average power at the circuit’s resonant frequency.

Strategy and Solution for (a)

The power factor at 60.0 Hz is found from

Equation:

$$\cos\,\phi = \frac{R}{Z}.$$

We know Z= 531 Ω from [link], so that

Equation:

$$\cos\phi = \frac{40.0\ \Omega}{531\ \Omega} = 0.0753 \text{ at } 60.0\text{ Hz.}$$

This small value indicates the voltage and current are significantly out of phase. In fact, the phase angle is

Equation:

$$\phi = \cos^{-1} 0.0753 = 85.7^{\text{o}} \text{ at } 60.0\text{ Hz.}$$

Discussion for (a)

The phase angle is close to 90º, consistent with the fact that the capacitor dominates the circuit at this low frequency (a pure *RC* circuit has its voltage and current 90º out of phase).

Strategy and Solution for (b)

The average power at 60.0 Hz is

Equation:

$$P_{\text{ave}} = I_{\text{rms}}V_{\text{rms}}\cos\phi.$$

I_{rms} was found to be 0.226 A in [link]. Entering the known values gives

Equation:

$$P_{\text{ave}} = (0.226\text{ A})(120\text{ V})(0.0753) = 2.04\text{ W at } 60.0\text{ Hz.}$$

Strategy and Solution for (c)

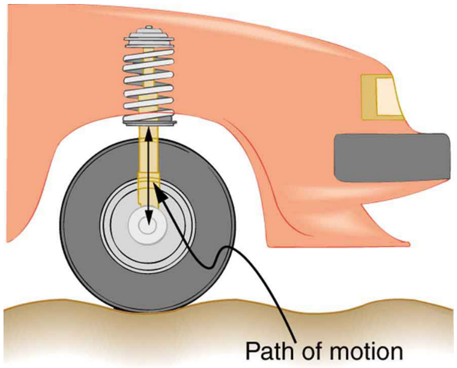
At the resonant frequency, we know $\cos\phi = 1$, and I_{rms} was found to be 6.00 A in [link]. Thus,

$P_{\text{ave}} = (3.00\text{ A})(120\text{ V})(1) = 360\text{ W}$ at resonance (1.30 kHz)

Discussion

Both the current and the power factor are greater at resonance, producing significantly greater power than at higher and lower frequencies.

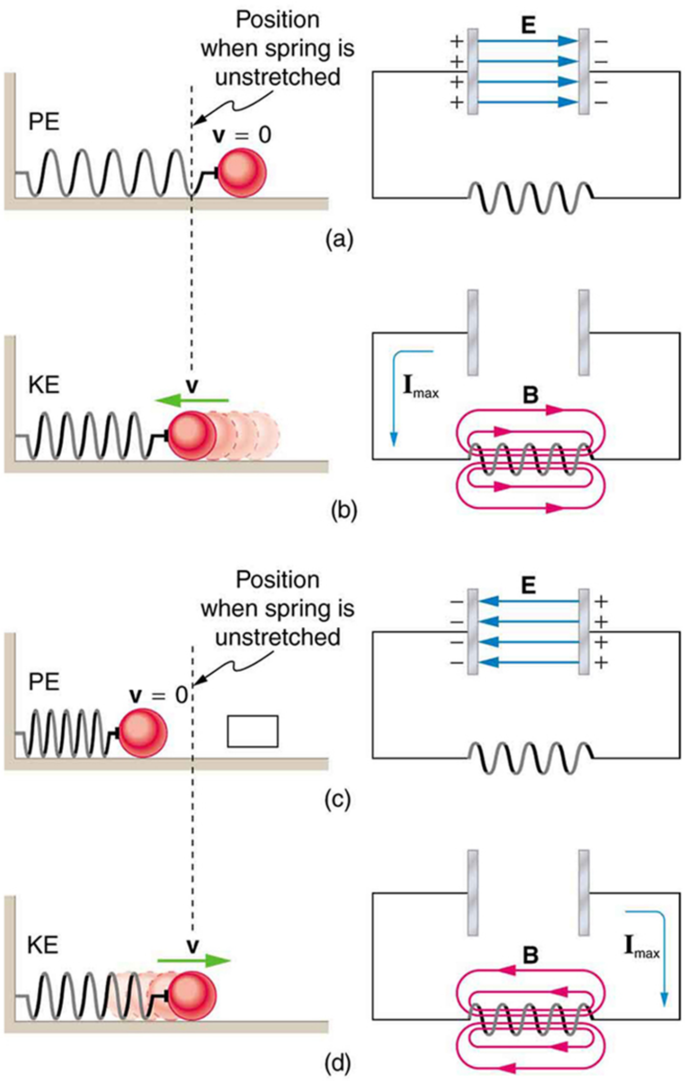
Power delivered to an *RLC* series AC circuit is dissipated by the resistance alone. The inductor and capacitor have energy input and output but do not dissipate it out of the circuit. Rather they transfer energy back and forth to one another, with the resistor dissipating exactly what the voltage source puts into the circuit. This assumes no significant electromagnetic radiation from the inductor and capacitor, such as radio waves. Such radiation can happen and may even be desired, as we will see in the next chapter on electromagnetic radiation, but it can also be suppressed as is the case in this chapter. The circuit is analogous to the wheel of a car driven over a corrugated road as shown in [link]. The regularly spaced bumps in the road are analogous to the voltage source, driving the wheel up and down. The shock absorber is analogous to the resistance damping and limiting the amplitude of the oscillation. Energy within the system goes back and forth between kinetic (analogous to maximum current, and energy stored in an inductor) and potential energy stored in the car spring (analogous to no current, and energy stored in the electric field of a capacitor). The amplitude of the wheels' motion is a maximum if the bumps in the road are hit at the resonant frequency.



The forced but damped motion of the wheel on the car spring is analogous to an *RLC* series AC circuit. The shock absorber damps the motion and dissipates energy, analogous to the resistance in an *RLC* circuit. The mass

and spring determine
the resonant
frequency.

A pure *LC* circuit with negligible resistance oscillates at f_0, the same resonant frequency as an *RLC* circuit. It can serve as a frequency standard or clock circuit—for example, in a digital wristwatch. With a very small resistance, only a very small energy input is necessary to maintain the oscillations. The circuit is analogous to a car with no shock absorbers. Once it starts oscillating, it continues at its natural frequency for some time. [link] shows the analogy between an *LC* circuit and a mass on a spring.


Position when spring is unstretched
PE
v = 0
E
+
−
(a)
KE
v
I_max
B
(b)
Position when spring is unstretched
PE
v = 0
E
−
+
(c)
KE
v
B
I_max
(d)

An *LC* circuit is analogous to a mass oscillating on a spring with no friction and no driving force. Energy moves back and forth between the inductor and capacitor, just as it moves from kinetic to potential in the mass-spring system.

Note:
PhET Explorations: Circuit Construction Kit (AC+DC), Virtual Lab
Build circuits with capacitors, inductors, resistors and AC or DC voltage sources, and inspect them using lab instruments such as voltmeters and ammeters.
https://phet.colorado.edu/sims/html/circuit-construction-kit-dc/latest/circuit-construction-kit-dc_en.html

Section Summary

- The AC analogy to resistance is impedance Z, the combined effect of resistors, inductors, and capacitors, defined by the AC version of Ohm's law:
  Equation:
  $$I_0 = \frac{V_0}{Z} \text{ or } I_{\text{rms}} = \frac{V_{\text{rms}}}{Z},$$
  where I_0 is the peak current and V_0 is the peak source voltage.
- Impedance has units of ohms and is given by $Z = \sqrt{R^2 + (X_L - X_C)^2}$.
- The resonant frequency f_0, at which $X_L = X_C$, is
  Equation:
  $$f_0 = \frac{1}{2\pi\sqrt{LC}}.$$
- In an AC circuit, there is a phase angle ϕ between source voltage V and the current I, which can be found from
  Equation:
  $$\cos\phi = \frac{R}{Z},$$
- $\phi = 0^\circ$ for a purely resistive circuit or an *RLC* circuit at resonance.

- The average power delivered to an *RLC* circuit is affected by the phase angle and is given by
  Equation:
  $$P_{\text{ave}} = I_{\text{rms}}V_{\text{rms}} \cos \phi,$$
  $\cos \phi$ is called the power factor, which ranges from 0 to 1.

Conceptual Questions

Exercise:

Problem:

Does the resonant frequency of an AC circuit depend on the peak voltage of the AC source? Explain why or why not.

Exercise:

Problem:

Suppose you have a motor with a power factor significantly less than 1. Explain why it would be better to improve the power factor as a method of improving the motor’s output, rather than to increase the voltage input.

Problems & Exercises

Exercise:

Problem:

An *RL* circuit consists of a 40.0 Ω resistor and a 3.00 mH inductor. (a) Find its impedance Z at 60.0 Hz and 10.0 kHz. (b) Compare these values of Z with those found in [link] in which there was also a capacitor.

Solution:

(a) $40.02\ \Omega$ at 60.0 Hz, $193\ \Omega$ at 10.0 kHz

(b) At 60 Hz, with a capacitor, $\mathrm{Z}=531\ \Omega$, over 13 times as high as without the capacitor. The capacitor makes a large difference at low frequencies. At 10 kHz, with a capacitor $\mathrm{Z}=190\ \Omega$, about the same as without the capacitor. The capacitor has a smaller effect at high frequencies.

Exercise:

Problem:

An *RC* circuit consists of a $40.0\ \Omega$ resistor and a $5.00\ \mu\mathrm{F}$ capacitor. (a) Find its impedance at 60.0 Hz and 10.0 kHz. (b) Compare these values of Z with those found in [link], in which there was also an inductor.

Exercise:

Problem:

An *LC* circuit consists of a $3.00\ \mathrm{mH}$ inductor and a $5.00\ \mu F$ capacitor. (a) Find its impedance at 60.0 Hz and 10.0 kHz. (b) Compare these values of Z with those found in [link] in which there was also a resistor.

Solution:

(a) $529\ \Omega$ at 60.0 Hz, $185\ \Omega$ at 10.0 kHz

(b) These values are close to those obtained in [link] because at low frequency the capacitor dominates and at high frequency the inductor dominates. So in both cases the resistor makes little contribution to the total impedance.

Exercise:

Problem:

What is the resonant frequency of a 0.500 mH inductor connected to a $40.0\ \mu\mathrm{F}$ capacitor?

Exercise:

Problem:

To receive AM radio, you want an *RLC* circuit that can be made to resonate at any frequency between 500 and 1650 kHz. This is accomplished with a fixed $1.00\ \mu\text{H}$ inductor connected to a variable capacitor. What range of capacitance is needed?

Solution:

9.30 nF to 101 nF

Exercise:

Problem:

Suppose you have a supply of inductors ranging from 1.00 nH to 10.0 H, and capacitors ranging from 1.00 pF to 0.100 F. What is the range of resonant frequencies that can be achieved from combinations of a single inductor and a single capacitor?

Exercise:

Problem:

What capacitance do you need to produce a resonant frequency of 1.00 GHz, when using an 8.00 nH inductor?

Solution:

3.17 pF

Exercise:

Problem:

What inductance do you need to produce a resonant frequency of 60.0 Hz, when using a $2.00\ \mu\text{F}$ capacitor?

Exercise:

Problem:

The lowest frequency in the FM radio band is 88.0 MHz. (a) What inductance is needed to produce this resonant frequency if it is connected to a 2.50 pF capacitor? (b) The capacitor is variable, to allow the resonant frequency to be adjusted to as high as 108 MHz. What must the capacitance be at this frequency?

Solution:

(a) $1.31\ \mu\text{H}$

(b) 1.66 pF

Exercise:

Problem:

An *RLC* series circuit has a $2.50\ \Omega$ resistor, a $100\ \mu\text{H}$ inductor, and an $80.0\ \mu\text{F}$ capacitor.(a) Find the circuit’s impedance at 120 Hz. (b) Find the circuit’s impedance at 5.00 kHz. (c) If the voltage source has $V_{\text{rms}} = 5.60\ \text{V}$, what is I_{rms} at each frequency? (d) What is the resonant frequency of the circuit? (e) What is I_{rms} at resonance?

Exercise:

Problem:

An *RLC* series circuit has a $1.00\ \text{k}\Omega$ resistor, a $150\ \mu\text{H}$ inductor, and a 25.0 nF capacitor. (a) Find the circuit’s impedance at 500 Hz. (b) Find the circuit’s impedance at 7.50 kHz. (c) If the voltage source has $V_{\text{rms}} = 408\ \text{V}$, what is I_{rms} at each frequency? (d) What is the resonant frequency of the circuit? (e) What is I_{rms} at resonance?

Solution:

(a) $12.8\ \text{k}\Omega$

(b) $1.31\ \text{k}\Omega$

(c) 31.9 mA at 500 Hz, 312 mA at 7.50 kHz

(d) 82.2 kHz

(e) 0.408 A

Exercise:

Problem:

An *RLC* series circuit has a $2.50\ \Omega$ resistor, a $100\ \mu\text{H}$ inductor, and an $80.0\ \mu\text{F}$ capacitor. (a) Find the power factor at $f = 120\ \text{Hz}$. (b) What is the phase angle at 120 Hz? (c) What is the average power at 120 Hz? (d) Find the average power at the circuit's resonant frequency.

Exercise:

Problem:

An *RLC* series circuit has a $1.00\ \text{k}\Omega$ resistor, a $150\ \mu\text{H}$ inductor, and a 25.0 nF capacitor. (a) Find the power factor at $f = 7.50\ \text{Hz}$. (b) What is the phase angle at this frequency? (c) What is the average power at this frequency? (d) Find the average power at the circuit's resonant frequency.

Solution:

(a) 0.159

(b) 80.9º

(c) 26.4 W

(d) 166 W

Exercise:

Problem:

An *RLC* series circuit has a $200\ \Omega$ resistor and a 25.0 mH inductor. At 8000 Hz, the phase angle is 45.0°. (a) What is the impedance? (b) Find the circuit’s capacitance. (c) If $V_{\text{rms}} = 408\text{ V}$ is applied, what is the average power supplied?

Exercise:

Problem: Referring to [link], find the average power at 10.0 kHz.

Solution:

16.0 W

Glossary

impedance
: the AC analogue to resistance in a DC circuit; it is the combined effect of resistance, inductive reactance, and capacitive reactance in the form $Z = \sqrt{R^2 + (X_L - X_C)^2}$

resonant frequency
: the frequency at which the impedance in a circuit is at a minimum, and also the frequency at which the circuit would oscillate if not driven by a voltage source; calculated by $f_0 = \frac{1}{2\pi\sqrt{\text{LC}}}$

phase angle
: denoted by ϕ, the amount by which the voltage and current are out of phase with each other in a circuit

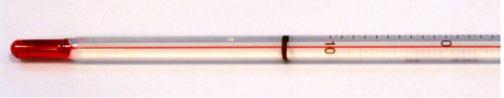
power factor
: the amount by which the power delivered in the circuit is less than the theoretical maximum of the circuit due to voltage and current being out of phase; calculated by $\cos\phi$

Introduction to Temperature, Kinetic Theory, and the Gas Laws class="introduction"

The welder’s gloves and helmet protect him from the electric arc that transfers enough thermal energy to melt the rod, spray sparks, and burn the retina of an unprotected eye. The thermal energy can be felt on exposed skin a few meters away, and its light can be seen for kilometers. (credit: Kevin S. O’Brien/U.S. Navy)

Heat is something familiar to each of us. We feel the warmth of the summer Sun, the chill of a clear summer night, the heat of coffee after a winter stroll, and the cooling effect of our sweat. Heat transfer is maintained by temperature differences. Manifestations of **heat transfer**—the movement of heat energy from one place or material to another—are apparent throughout the universe. Heat from beneath Earth’s surface is brought to the surface in flows of incandescent lava. The Sun warms Earth’s surface and is the source of much of the energy we find on it. Rising levels of atmospheric carbon dioxide threaten to trap more of the Sun’s energy, perhaps fundamentally altering the ecosphere. In space, supernovas explode, briefly radiating more heat than an entire galaxy does.

What is heat? How do we define it? How is it related to temperature? What are heat’s effects? How is it related to other forms of energy and to work? We will find that, in spite of the richness of the phenomena, there is a small set of underlying physical principles that unite the subjects and tie them to other fields.



In a typical thermometer like this one, the alcohol, with a red dye, expands more rapidly than the glass containing it. When the thermometer’s temperature increases, the liquid from the bulb is forced into the narrow tube, producing a large change in the length of the column for a small change in temperature. (credit: Chemical Engineer, Wikimedia Commons)

Temperature

- Define temperature.
- Convert temperatures between the Celsius, Fahrenheit, and Kelvin scales.
- Define thermal equilibrium.
- State the zeroth law of thermodynamics.

The concept of temperature has evolved from the common concepts of hot and cold. Human perception of what feels hot or cold is a relative one. For example, if you place one hand in hot water and the other in cold water, and then place both hands in tepid water, the tepid water will feel cool to the hand that was in hot water, and warm to the one that was in cold water. The scientific definition of temperature is less ambiguous than your senses of hot and cold. **Temperature** is operationally defined to be what we measure with a thermometer. (Many physical quantities are defined solely in terms of how they are measured. We shall see later how temperature is related to the kinetic energies of atoms and molecules, a more physical explanation.) Two accurate thermometers, one placed in hot water and the other in cold water, will show the hot water to have a higher temperature. If they are then placed in the tepid water, both will give identical readings (within measurement uncertainties). In this section, we discuss temperature, its measurement by thermometers, and its relationship to thermal equilibrium. Again, temperature is the quantity measured by a thermometer.

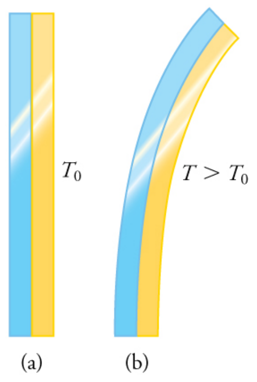
Note:

Misconception Alert: Human Perception vs. Reality

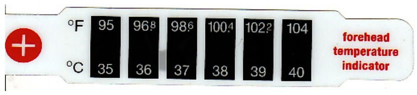
On a cold winter morning, the wood on a porch feels warmer than the metal of your bike. The wood and bicycle are in thermal equilibrium with the outside air, and are thus the same temperature. They *feel* different because of the difference in the way that they conduct heat away from your skin. The metal conducts heat away from your body faster than the wood does (see more about conductivity in Conduction). This is just one example demonstrating that the human sense of hot and cold is not determined by temperature alone.

Another factor that affects our perception of temperature is humidity. Most people feel much hotter on hot, humid days than on hot, dry days. This is because on humid days, sweat does not evaporate from the skin as efficiently as it does on dry days. It is the evaporation of sweat (or water from a sprinkler or pool) that cools us off.

Any physical property that depends on temperature, and whose response to temperature is reproducible, can be used as the basis of a thermometer. Because many physical properties depend on temperature, the variety of thermometers is remarkable. For example, volume increases with temperature for most substances. This property is the basis for the common alcohol thermometer, the old mercury thermometer, and the bimetallic strip ([link]). Other properties used to measure temperature include electrical resistance and color, as shown in [link], and the emission of infrared radiation, as shown in [link].



The curvature of a bimetallic strip depends on temperature. (a) The strip is straight at the starting temperature, where its two components have the same length. (b) At a higher temperature, this strip bends to the right, because the metal on the left has expanded more than the metal on the right.


°F
95
96.8
98.6
100.4
102.2
104
°C
35
36
37
38
39
40
forehead temperature indicator

Each of the six squares on this plastic (liquid crystal) thermometer contains a film of a different heat-sensitive liquid crystal material. Below 95ºF, all six squares are black. When the plastic thermometer is exposed to temperature that increases to 95ºF, the first liquid crystal square changes color. When the temperature increases above 96.8ºF the second liquid crystal square also changes color, and so forth. (credit: Arkrishna, Wikimedia Commons)

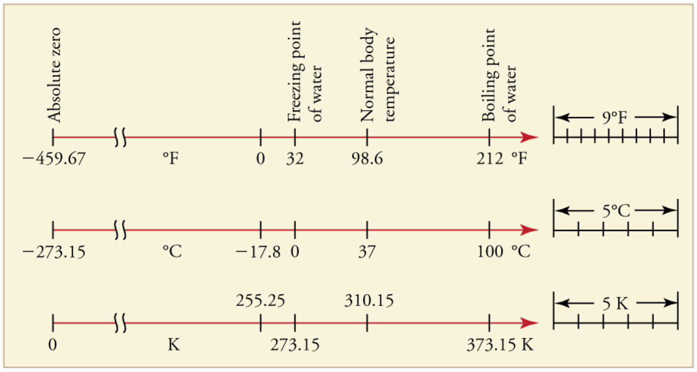
Fireman Jason Ormand uses a pyrometer to check the temperature of an aircraft carrier’s ventilation system. Infrared radiation (whose emission varies with temperature) from the vent is measured and a temperature readout is quickly produced. Infrared measurements are also frequently used as a measure of body temperature. These modern thermometers, placed in the ear canal, are more accurate than alcohol thermometers placed under the tongue or in the armpit. (credit: Lamel J. Hinton/U.S. Navy)

Temperature Scales

Thermometers are used to measure temperature according to well-defined scales of measurement, which use pre-defined reference points to help compare quantities. The three most common temperature scales are the Fahrenheit, Celsius, and Kelvin scales. A temperature scale can be created by identifying two easily reproducible temperatures. The freezing and boiling temperatures of water at standard atmospheric pressure are commonly used.

The **Celsius** scale (which replaced the slightly different *centigrade* scale) has the freezing point of water at $0^{\circ}\mathrm{C}$ and the boiling point at $100^{\circ}\mathrm{C}$. Its unit is the **degree Celsius**$(^{\circ}\mathrm{C})$. On the **Fahrenheit** scale (still the most frequently used in the United States), the freezing point of water is at $32^{\circ}\mathrm{F}$ and the boiling point is at $212^{\circ}\mathrm{F}$. The unit of temperature on this scale is the **degree Fahrenheit**$(^{\circ}\mathrm{F})$. Note that a temperature difference of one degree Celsius is greater than a temperature difference of one degree Fahrenheit. Only 100 Celsius degrees span the same range as 180 Fahrenheit degrees, thus one degree on the Celsius scale is 1.8 times larger than one degree on the Fahrenheit scale $180/100 = 9/5$.

The **Kelvin** scale is the temperature scale that is commonly used in science. It is an *absolute temperature* scale defined to have 0 K at the lowest possible temperature, called **absolute zero**. The official temperature unit on this scale is the *kelvin*, which is abbreviated K, and is not accompanied by a degree sign. The freezing and boiling points of water are 273.15 K and 373.15 K, respectively. Thus, the magnitude of temperature differences is the same in units of kelvins and degrees Celsius. Unlike other temperature scales, the Kelvin scale is an absolute scale. It is used extensively in scientific work because a number of physical quantities, such as the volume of an ideal gas, are directly related to absolute temperature. The kelvin is the SI unit used in scientific work.


Absolute zero
Freezing point of water
Normal body temperature
Boiling point of water
9°F
−459.67
°F
0
32
98.6
212 °F
5°C
−273.15
°C
−17.8
0
37
100 °C
255.25
310.15
5 K
0
K
273.15
373.15 K

Relationships between the Fahrenheit, Celsius, and Kelvin temperature scales, rounded to the nearest degree. The relative sizes of the scales are also shown.

The relationships between the three common temperature scales is shown in [link]. Temperatures on these scales can be converted using the equations in [link].

To convert from . . .	**Use this equation . . .**	**Also written as . . .**

To convert from . . .	**Use this equation . . .**	**Also written as . . .**
Celsius to Fahrenheit	$T(^\mathrm{o}\mathrm{F}) = \frac{9}{5}T(^\mathrm{o}\mathrm{C}) + 32$	$T_{^\mathrm{o}\mathrm{F}} = \frac{9}{5}T_{^\mathrm{o}\mathrm{C}} + 32$
Fahrenheit to Celsius	$T(^\mathrm{o}\mathrm{C}) = \frac{5}{9}(T(^\mathrm{o}\mathrm{F}) - 32)$	$T_{^\mathrm{o}\mathrm{C}} = \frac{5}{9}\left(T_{^\mathrm{o}\mathrm{F}} - 32\right)$
Celsius to Kelvin	$T(\mathrm{K}) = T(^\mathrm{o}\mathrm{C}) + 273.15$	$T_\mathrm{K} = T_{^\mathrm{o}\mathrm{C}} + 273.15$
Kelvin to Celsius	$T(^\mathrm{o}\mathrm{C}) = T(\mathrm{K}) - 273.15$	$T_{^\mathrm{o}\mathrm{C}} = T_\mathrm{K} - 273.15$
Fahrenheit to Kelvin	$T(\mathrm{K}) = \frac{5}{9}(T(^\mathrm{o}\mathrm{F}) - 32) + 273.15$	$T_\mathrm{K} = \frac{5}{9}\left(T_{^\mathrm{o}\mathrm{F}} - 32\right) + 273.15$
Kelvin to Fahrenheit	$T(^\mathrm{o}\mathrm{F}) = \frac{9}{5}(T(\mathrm{K}) - 273.15) + 32$	$T_{^\mathrm{o}\mathrm{F}} = \frac{9}{5}(T_\mathrm{K} - 273.15) + 32$

Temperature Conversions

Notice that the conversions between Fahrenheit and Kelvin look quite complicated. In fact, they are simple combinations of the conversions between Fahrenheit and Celsius, and the conversions between Celsius and Kelvin.

Example:
Converting between Temperature Scales: Room Temperature
“Room temperature” is generally defined to be $25^\mathrm{o}\mathrm{C}$. (a) What is room temperature in ºF? (b) What is it in K?
Strategy
To answer these questions, all we need to do is choose the correct conversion equations and plug in the known values.

Solution for (a)

1. Choose the right equation. To convert from ºC to ºF, use the equation

Equation:

$$T_{°\text{F}} = \frac{9}{5}T_{°\text{C}} + 32.$$

2. Plug the known value into the equation and solve:

Equation:

$$T_{°\text{F}} = \frac{9}{5}25°\text{C} + 32 = 77°\text{F}.$$

Solution for (b)

1. Choose the right equation. To convert from ºC to K, use the equation

Equation:

$$T_{\text{K}} = T_{°\text{C}} + 273.15.$$

2. Plug the known value into the equation and solve:

Equation:

$$T_{\text{K}} = 25°\text{C} + 273.15 = 298\,\text{K}.$$

Example:

Converting between Temperature Scales: the Reaumur Scale

The Reaumur scale is a temperature scale that was used widely in Europe in the 18th and 19th centuries. On the Reaumur temperature scale, the freezing point of water is $0°\text{R}$ and the boiling temperature is $80°\text{R}$. If “room temperature” is $25°\text{C}$ on the Celsius scale, what is it on the Reaumur scale?

Strategy

To answer this question, we must compare the Reaumur scale to the Celsius scale. The difference between the freezing point and boiling point of water on the Reaumur scale is $80°\text{R}$. On the Celsius scale it is $100°\text{C}$. Therefore $100°\text{C} = 80°\text{R}$. Both scales start at $0°$ for freezing, so we can derive a simple formula to convert between temperatures on the two scales.

Solution

1. Derive a formula to convert from one scale to the other:

Equation:

$$T_{°\text{R}} = \frac{0.8°\text{R}}{°\text{C}} \times T_{°\text{C}}.$$

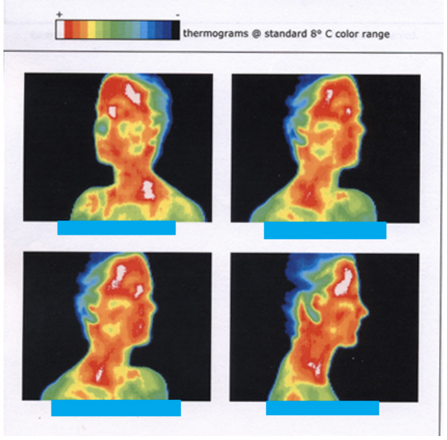
2. Plug the known value into the equation and solve:

Equation:

$$T_{^\circ\text{R}} = \frac{0.8^\circ\text{R}}{^\circ\text{C}} \times 25^\circ\text{C} = 20^\circ\text{R}.$$

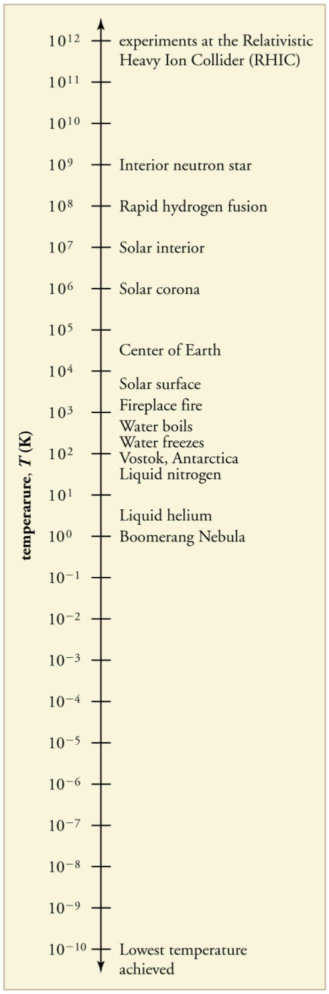
Temperature Ranges in the Universe

[link] shows the wide range of temperatures found in the universe. Human beings have been known to survive with body temperatures within a small range, from 24°C to 44°C $(75^\circ\text{F}$ to $111^\circ\text{F})$. The average normal body temperature is usually given as 37.0°C (98.6°F), and variations in this temperature can indicate a medical condition: a fever, an infection, a tumor, or circulatory problems (see [link]).



This image of radiation from a person’s body (an infrared thermograph) shows the location of temperature abnormalities in the upper body. Dark blue corresponds to cold areas and red to white corresponds to hot areas. An elevated temperature might be an indication of malignant tissue (a cancerous tumor in the breast, for example), while a depressed temperature might be due to a decline in blood flow from a clot. In this case, the abnormalities are caused by a condition called hyperhidrosis. (credit: Porcelina81, Wikimedia Commons)

The lowest temperatures ever recorded have been measured during laboratory experiments: $4.5 \times 10^{-10}\ \text{K}$ at the Massachusetts Institute of Technology (USA), and $1.0 \times 10^{-10}\ \text{K}$ at Helsinki University of Technology (Finland). In comparison, the coldest recorded place on Earth's surface is Vostok, Antarctica at 183 K $\left(-89^{\text{o}}\text{C}\right)$, and the coldest place (outside the lab) known in the universe is the Boomerang Nebula, with a temperature of 1 K.

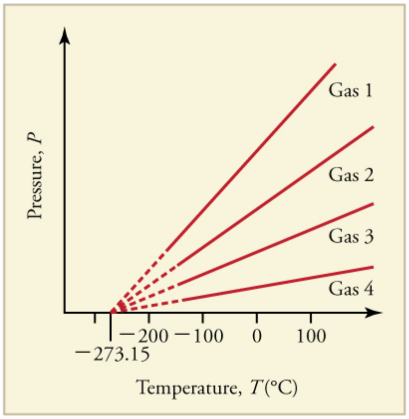


Each increment on this logarithmic scale indicates an increase by a factor of ten, and thus illustrates the tremendous range of temperatures in nature. Note that zero on a logarithmic scale would occur off the bottom of the page at infinity.

Note:

Making Connections: Absolute Zero

What is absolute zero? Absolute zero is the temperature at which all molecular motion has ceased. The concept of absolute zero arises from the behavior of gases. [link] shows how the pressure of gases at a constant volume decreases as temperature decreases. Various scientists have noted that the pressures of gases extrapolate to zero at the same temperature, -273.15°C. This extrapolation implies that there is a lowest temperature. This temperature is called *absolute zero*. Today we know that most gases first liquefy and then freeze, and it is not actually possible to reach absolute zero. The numerical value of absolute zero temperature is -273.15°C or 0 K.



Graph of pressure versus temperature for various gases kept at a constant volume. Note that all of the graphs extrapolate to zero pressure at the same temperature.

Thermal Equilibrium and the Zeroth Law of Thermodynamics

Thermometers actually take their *own* temperature, not the temperature of the object they are measuring. This raises the question of how we can be certain that a thermometer measures the temperature of the object with which it is in contact. It is based on the fact that any two systems placed in *thermal contact* (meaning heat transfer can occur between them) will reach the same temperature. That is, heat will flow from the hotter object to the cooler one until they have exactly the same temperature. The objects are then in **thermal equilibrium**, and no further changes will occur. The systems interact and change because their temperatures differ, and the changes stop once their temperatures are the same. Thus, if enough time is allowed for this transfer of heat to run its course, the temperature a thermometer registers *does* represent the system with which it is in thermal equilibrium. Thermal equilibrium is established when two bodies are in contact with each other and can freely exchange energy.

Furthermore, experimentation has shown that if two systems, A and B, are in thermal equilibrium with each another, and B is in thermal equilibrium with a third system C, then A is also in thermal equilibrium with C. This conclusion may seem obvious, because all three have the same temperature, but it is basic to thermodynamics. It is called the **zeroth law of thermodynamics**.

Note:
The Zeroth Law of Thermodynamics
If two systems, A and B, are in thermal equilibrium with each other, and B is in thermal equilibrium with a third system, C, then A is also in thermal equilibrium with C.

This law was postulated in the 1930s, after the first and second laws of thermodynamics had been developed and named. It is called the *zeroth law* because it comes logically before the first and second laws (discussed in Thermodynamics). An example of this law in action is seen in babies in incubators: babies in incubators normally have very few clothes on, so to an observer they look as if they may not be warm enough. However, the temperature of the air, the cot, and the baby is the same, because they are in thermal equilibrium, which is accomplished by maintaining air temperature to keep the baby comfortable.

Exercise:
Check Your Understanding

Problem: Does the temperature of a body depend on its size?

Solution:

No, the system can be divided into smaller parts each of which is at the same temperature. We say that the temperature is an *intensive* quantity. Intensive quantities are independent of size.

Section Summary

- Temperature is the quantity measured by a thermometer.
- Temperature is related to the average kinetic energy of atoms and molecules in a system.
- Absolute zero is the temperature at which there is no molecular motion.
- There are three main temperature scales: Celsius, Fahrenheit, and Kelvin.
- Temperatures on one scale can be converted to temperatures on another scale using the following equations:
  Equation:
  $$T_{\text{°F}} = \frac{9}{5} T_{\text{°C}} + 32$$
  Equation:
  $$T_{\text{°C}} = \frac{5}{9} \left(T_{\text{°F}} - 32\right)$$
  Equation:
  $$T_{\text{K}} = T_{\text{°C}} + 273.15$$
  Equation:
  $$T_{\text{°C}} = T_{\text{K}} - 273.15$$
- Systems are in thermal equilibrium when they have the same temperature.
- Thermal equilibrium occurs when two bodies are in contact with each other and can freely exchange energy.
- The zeroth law of thermodynamics states that when two systems, A and B, are in thermal equilibrium with each other, and B is in thermal equilibrium with a third system, C, then A is also in thermal equilibrium with C.

Conceptual Questions

Exercise:

Problem: What does it mean to say that two systems are in thermal equilibrium?

Exercise:

Problem:

Give an example of a physical property that varies with temperature and describe how it is used to measure temperature.

Exercise:

Problem:

When a cold alcohol thermometer is placed in a hot liquid, the column of alcohol goes *down* slightly before going up. Explain why.

Exercise:

Problem:

If you add boiling water to a cup at room temperature, what would you expect the final equilibrium temperature of the unit to be? You will need to include the surroundings as part of the system. Consider the zeroth law of thermodynamics.

Problems & Exercises

Exercise:

Problem: What is the Fahrenheit temperature of a person with a 39.0°C fever?

Solution:

102°F

Exercise:

Problem:

Frost damage to most plants occurs at temperatures of 28.0°F or lower. What is this temperature on the Kelvin scale?

Exercise:

Problem:

To conserve energy, room temperatures are kept at 68.0°F in the winter and 78.0°F in the summer. What are these temperatures on the Celsius scale?

Solution:

20.0°C and 25.6°C

Exercise:

Problem:

A tungsten light bulb filament may operate at 2900 K. What is its Fahrenheit temperature? What is this on the Celsius scale?

Exercise:

Problem:

The surface temperature of the Sun is about 5750 K. What is this temperature on the Fahrenheit scale?

Solution:

9890°F

Exercise:

Problem:

One of the hottest temperatures ever recorded on the surface of Earth was 134°F in Death Valley, CA. What is this temperature in Celsius degrees? What is this temperature in Kelvin?

Exercise:

Problem:

(a) Suppose a cold front blows into your locale and drops the temperature by 40.0 Fahrenheit degrees. How many degrees Celsius does the temperature decrease when there is a 40.0°F decrease in temperature? (b) Show that any change in temperature in Fahrenheit degrees is nine-fifths the change in Celsius degrees.

Solution:

(a) 22.2°C

(b)

$$\begin{aligned}\Delta T(^\circ\text{F}) &= T_2(^\circ\text{F}) - T_1(^\circ\text{F}) \\ &= \tfrac{9}{5}T_2(^\circ\text{C}) + 32.0^\circ - \left(\tfrac{9}{5}T_1(^\circ\text{C}) + 32.0^\circ\right) \\ &= \tfrac{9}{5}\left(T_2(^\circ\text{C}) - T_1(^\circ\text{C})\right) = \tfrac{9}{5}\Delta T(^\circ\text{C})\end{aligned}$$

Exercise:

Problem:

(a) At what temperature do the Fahrenheit and Celsius scales have the same numerical value? (b) At what temperature do the Fahrenheit and Kelvin scales have the same numerical value?

Glossary

temperature	the quantity measured by a thermometer
Celsius scale	temperature scale in which the freezing point of water is 0°C and the boiling point of water is 100°C
degree Celsius	unit on the Celsius temperature scale
Fahrenheit scale	temperature scale in which the freezing point of water is 32°F and the boiling point of water is 212°F
degree Fahrenheit	unit on the Fahrenheit temperature scale
Kelvin scale	temperature scale in which 0 K is the lowest possible temperature, representing absolute zero
absolute zero	the lowest possible temperature; the temperature at which all molecular motion ceases
thermal equilibrium	the condition in which heat no longer flows between two objects that are in contact; the two objects have the same temperature
zeroth law of thermodynamics	law that states that if two objects are in thermal equilibrium, and a third object is in thermal equilibrium with one of those objects, it is also in thermal equilibrium with the other object

Thermal Expansion of Solids and Liquids

- Define and describe thermal expansion.
- Calculate the linear expansion of an object given its initial length, change in temperature, and coefficient of linear expansion.
- Calculate the volume expansion of an object given its initial volume, change in temperature, and coefficient of volume expansion.
- Calculate thermal stress on an object given its original volume, temperature change, volume change, and bulk modulus.

Thermal expansion joints like these in the Auckland Harbour Bridge in New Zealand allow bridges to change length without buckling. (credit: Ingolfson, Wikimedia Commons)

The expansion of alcohol in a thermometer is one of many commonly encountered examples of **thermal expansion**, the change in size or volume of a given mass with temperature. Hot air rises because its volume increases, which causes the hot air’s density to be smaller than the density of surrounding air, causing a buoyant (upward) force on the hot air. The same happens in all liquids and gases, driving natural heat transfer upwards in homes, oceans, and weather systems. Solids also undergo thermal expansion. Railroad tracks and bridges, for example, have expansion joints to allow them to freely expand and contract with temperature changes.

What are the basic properties of thermal expansion? First, thermal expansion is clearly related to temperature change. The greater the temperature change, the more a bimetallic strip will bend. Second, it depends on the material. In a thermometer, for example, the expansion of alcohol is much greater than the expansion of the glass containing it.

What is the underlying cause of thermal expansion? As is discussed in Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature, an increase in temperature implies an increase in the kinetic energy of the individual atoms. In a solid, unlike in a gas, the atoms or molecules are closely packed together, but their kinetic energy (in the form of small, rapid vibrations) pushes neighboring atoms or molecules apart from each other. This neighbor-to-neighbor pushing results in a slightly greater distance, on average, between neighbors, and adds up to a larger size for the whole body. For most substances under ordinary conditions, there is no preferred direction, and an increase in temperature will increase the solid’s size by a certain fraction in each dimension.

Note:

Linear Thermal Expansion—Thermal Expansion in One Dimension

The change in length ΔL is proportional to length L. The dependence of thermal expansion on temperature, substance, and length is summarized in the equation

Equation:

$$\Delta L = \alpha L \Delta T,$$

where ΔL is the change in length L, ΔT is the change in temperature, and α is the **coefficient of linear expansion**, which varies slightly with temperature.

[link] lists representative values of the coefficient of linear expansion, which may have units of $1/^{\text{o}}\text{C}$ or 1/K. Because the size of a kelvin and a degree Celsius are the same, both α and ΔT can be expressed in units of kelvins or degrees Celsius. The equation $\Delta L = \alpha L \Delta T$ is accurate for small changes in temperature and can be used for large changes in temperature if an average value of α is used.

Material	**Coefficient of linear expansion** $\alpha(1/^{\text{o}}\text{C})$	**Coefficient of volume expansion** $\beta(1/^{\text{o}}\text{C})$
Solids		
Aluminum	25×10^{-6}	75×10^{-6}
Brass	19×10^{-6}	56×10^{-6}
Copper	17×10^{-6}	51×10^{-6}

Material	**Coefficient of linear expansion** $\alpha(1/^{\text{o}}\text{C})$	**Coefficient of volume expansion** $\beta(1/^{\text{o}}\text{C})$
Gold	14×10^{-6}	42×10^{-6}
Iron or Steel	12×10^{-6}	35×10^{-6}
Invar (Nickel-iron alloy)	0.9×10^{-6}	2.7×10^{-6}
Lead	29×10^{-6}	87×10^{-6}
Silver	18×10^{-6}	54×10^{-6}
Glass (ordinary)	9×10^{-6}	27×10^{-6}
Glass (Pyrex®)	3×10^{-6}	9×10^{-6}
Quartz	0.4×10^{-6}	1×10^{-6}

Material	**Coefficient of linear expansion** $\alpha(1/^{o}\mathrm{C})$	**Coefficient of volume expansion** $\beta(1/^{o}\mathrm{C})$
Concrete, Brick	$\sim 12 \times 10^{-6}$	$\sim 36 \times 10^{-6}$
Marble (average)	7×10^{-6}	2.1×10^{-5}
Liquids		
Ether		1650×10^{-6}
Ethyl alcohol		1100×10^{-6}
Petrol		950×10^{-6}
Glycerin		500×10^{-6}
Mercury		180×10^{-6}

Material	**Coefficient of linear expansion** $\alpha(1/^\circ\text{C})$	**Coefficient of volume expansion** $\beta(1/^\circ\text{C})$
Water		210×10^{-6}
Gases		
Air and most other gases at atmospheric pressure		3400×10^{-6}

Thermal Expansion Coefficients at 20°C**[footnote]**
Values for liquids and gases are approximate.

Example:
Calculating Linear Thermal Expansion: The Golden Gate Bridge

The main span of San Francisco’s Golden Gate Bridge is 1275 m long at its coldest. The bridge is exposed to temperatures ranging from -15°C to 40°C. What is its change in length between these temperatures? Assume that the bridge is made entirely of steel.

Strategy

Use the equation for linear thermal expansion $\Delta L = \alpha L \Delta T$ to calculate the change in length , ΔL. Use the coefficient of linear expansion, α, for steel from [link], and note that the change in temperature, ΔT, is 55°C.

Solution

Plug all of the known values into the equation to solve for ΔL.

Equation:

$$\Delta L = \alpha L \Delta T = \left(\frac{12 \times 10^{-6}}{^\circ\text{C}}\right)(1275\ \text{m})(55^\circ\text{C}) = 0.84\ \text{m}.$$

Discussion
Although not large compared with the length of the bridge, this change in length is observable. It is generally spread over many expansion joints so that the expansion at each joint is small.

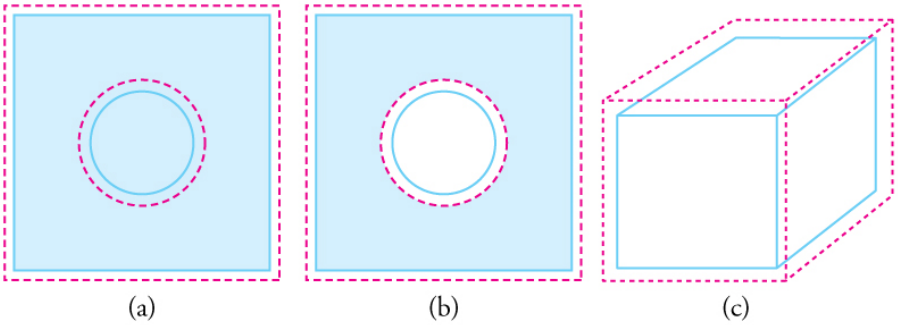
Thermal Expansion in Two and Three Dimensions

Objects expand in all dimensions, as illustrated in [link]. That is, their areas and volumes, as well as their lengths, increase with temperature. Holes also get larger with temperature. If you cut a hole in a metal plate, the remaining material will expand exactly as it would if the plug was still in place. The plug would get bigger, and so the hole must get bigger too. (Think of the ring of neighboring atoms or molecules on the wall of the hole as pushing each other farther apart as temperature increases. Obviously, the ring of neighbors must get slightly larger, so the hole gets slightly larger).

Note:
Thermal Expansion in Two Dimensions
For small temperature changes, the change in area ΔA is given by
Equation:

$$\Delta A = 2\alpha A \Delta T,$$

where ΔA is the change in area A, ΔT is the change in temperature, and α is the coefficient of linear expansion, which varies slightly with temperature.



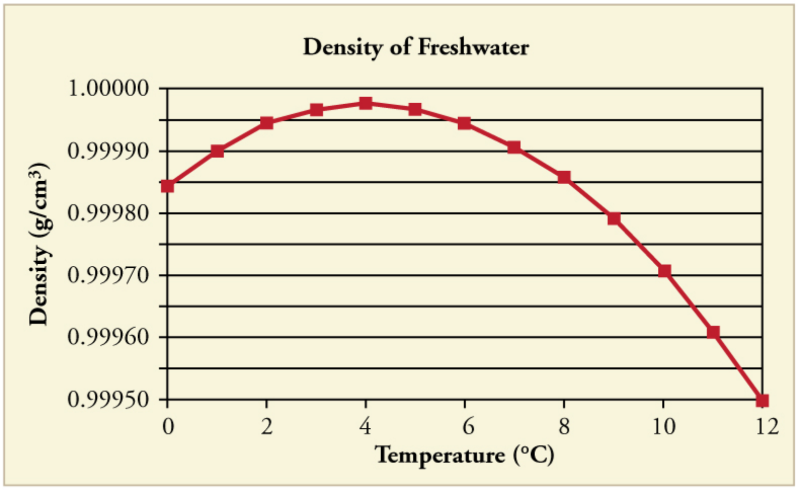
In general, objects expand in all directions as temperature increases. In these drawings, the original boundaries of the objects are shown with solid lines, and the expanded boundaries with dashed lines. (a) Area increases because both length and width increase. The area of a circular plug also increases. (b) If the plug is removed, the hole it leaves becomes larger with increasing temperature, just as if the expanding plug were still in place. (c) Volume also increases, because all three dimensions increase.

Note:
Thermal Expansion in Three Dimensions
The change in volume ΔV is very nearly $\Delta V = 3\alpha V \Delta T$. This equation is usually written as
Equation:

$$\Delta V = \beta V \Delta T,$$

where β is the **coefficient of volume expansion** and $\beta \approx 3\alpha$. Note that the values of β in [link] are almost exactly equal to 3α.

In general, objects will expand with increasing temperature. Water is the most important exception to this rule. Water expands with increasing temperature (its density *decreases*) when it is at temperatures greater than $4^\circ\text{C}(40^\circ\text{F})$. However, it expands with *decreasing* temperature when it is between $+4^\circ\text{C}$ and $0^\circ\text{C}(40^\circ\text{F}$ to $32^\circ\text{F})$. Water is densest at $+4^\circ\text{C}$. (See [link].) Perhaps the most striking effect of this phenomenon is the freezing of water in a pond. When water near the surface cools down to 4°C it is denser than the remaining water and thus will sink to the bottom. This "turnover" results in a layer of warmer water near the surface, which is then cooled. Eventually the pond has a uniform temperature of 4°C. If the temperature in the surface layer drops below 4°C, the water is less dense than the water below, and thus stays near the top. As a result, the pond surface can completely freeze over. The ice on top of liquid water provides an insulating layer from winter's harsh exterior air temperatures. Fish and other aquatic life can survive in 4°C water beneath ice, due to this unusual characteristic of water. It also produces circulation of water in the pond that is necessary for a healthy ecosystem of the body of water.



The density of water as a function of temperature. Note that the thermal expansion is actually very small. The maximum density at $+4^\circ\text{C}$ is only *0.0075%* greater than the density at 2°C, and *0.012%* greater than that at 0°C.

Note:

Making Connections: Real-World Connections—Filling the Tank

Differences in the thermal expansion of materials can lead to interesting effects at the gas station. One example is the dripping of gasoline from a freshly filled tank on a hot day. Gasoline starts out at the temperature of the ground under the gas station, which is cooler than the air temperature above. The gasoline cools the steel tank when it is filled. Both gasoline and steel tank expand as they warm to air temperature, but gasoline expands much more than steel, and so it may overflow.

This difference in expansion can also cause problems when interpreting the gasoline gauge. The actual amount (mass) of gasoline left in the tank when the gauge hits "empty" is a lot less in the summer than in the winter. The gasoline has the same volume as it does in the winter when the "add fuel" light goes on, but because the gasoline has expanded, there is less mass. If you are used to getting another 40 miles on "empty" in the winter, beware—you will probably run out much more quickly in the summer.

NORMAL
E
F
UNLEADED
FUEL ONLY
FUEL
FILL

Because the gas expands more than the gas tank with increasing temperature, you can't drive as many miles on "empty" in the summer as you can in the winter.

(credit: Hector Alejandro, Flickr)

Example:

Calculating Thermal Expansion: Gas vs. Gas Tank

Suppose your 60.0-L (15.9-gal) steel gasoline tank is full of gas, so both the tank and the gasoline have a temperature of 15.0°C. How much gasoline has spilled by the time they warm to 35.0°C?

Strategy

The tank and gasoline increase in volume, but the gasoline increases more, so the amount spilled is the difference in their volume changes. (The gasoline tank can be treated as solid steel.) We can use the equation for volume expansion to calculate the change in volume of the gasoline and of the tank.

Solution

1. Use the equation for volume expansion to calculate the increase in volume of the steel tank:

Equation:

$$\Delta V_{\text{s}} = \beta_{\text{s}} V_{\text{s}} \Delta T.$$

2. The increase in volume of the gasoline is given by this equation:

Equation:

$$\Delta V_{\text{gas}} = \beta_{\text{gas}} V_{\text{gas}} \Delta T.$$

3. Find the difference in volume to determine the amount spilled as

Equation:

$$V_{\text{spill}} = \Delta V_{\text{gas}} - \Delta V_{\text{s}}.$$

Alternatively, we can combine these three equations into a single equation. (Note that the original volumes are equal.)

Equation:

$$\begin{aligned} V_{\text{spill}} &= (\beta_{\text{gas}} - \beta_{\text{s}})V\Delta T \\ &= \left[(950 - 35) \times 10^{-6}/^{\circ}\text{C}\right](60.0\text{ L})(20.0^{\circ}\text{C}) \\ &= 1.10\text{ L.} \end{aligned}$$

Discussion

This amount is significant, particularly for a 60.0-L tank. The effect is so striking because the gasoline and steel expand quickly. The rate of change in thermal properties is discussed in Heat and Heat Transfer Methods.

If you try to cap the tank tightly to prevent overflow, you will find that it leaks anyway, either around the cap or by bursting the tank. Tightly constricting the expanding gas is equivalent to compressing it, and both liquids and solids resist being compressed with extremely large forces. To avoid rupturing rigid containers, these containers have air gaps, which allow them to expand and contract without stressing them.

Thermal Stress

Thermal stress is created by thermal expansion or contraction (see Elasticity: Stress and Strain for a discussion of stress and strain). Thermal stress can be destructive, such as when expanding gasoline ruptures a tank. It can also be useful, for example, when two parts are joined together by heating one in manufacturing, then slipping it over the other and allowing the combination to cool. Thermal stress can explain many phenomena, such as the weathering of rocks and pavement by the expansion of ice when it freezes.

Example:
Calculating Thermal Stress: Gas Pressure

What pressure would be created in the gasoline tank considered in [link], if the gasoline increases in temperature from 15.0°C to 35.0°C without being allowed to expand? Assume that the bulk modulus B for gasoline is $1.00 \times 10^{9}\ \text{N/m}^{2}$. (For more on bulk modulus, see Elasticity: Stress and Strain.)

Strategy

To solve this problem, we must use the following equation, which relates a change in volume ΔV to pressure:

Equation:

$$\Delta V = \frac{1}{B}\frac{F}{A}V_0,$$

where F/A is pressure, V_0 is the original volume, and B is the bulk modulus of the material involved. We will use the amount spilled in [link] as the change in volume, ΔV.

Solution

1. Rearrange the equation for calculating pressure:

Equation:

$$P = \frac{F}{A} = \frac{\Delta V}{V_0}B.$$

2. Insert the known values. The bulk modulus for gasoline is $B = 1.00 \times 10^9 \text{ N/m}^2$. In the previous example, the change in volume $\Delta V = 1.10 \text{ L}$ is the amount that would spill. Here, $V_0 = 60.0 \text{ L}$ is the original volume of the gasoline. Substituting these values into the equation, we obtain

Equation:

$$P = \frac{1.10 \text{ L}}{60.0 \text{ L}}\left(1.00 \times 10^9 \text{ Pa}\right) = 1.83 \times 10^7 \text{ Pa}.$$

Discussion

This pressure is about 2500 lb/in^2, *much* more than a gasoline tank can handle.

Forces and pressures created by thermal stress are typically as great as that in the example above. Railroad tracks and roadways can buckle on hot days if they lack sufficient expansion joints. (See [link].) Power lines sag more in the summer than in the winter, and will snap in cold weather if there is insufficient slack. Cracks open and close in plaster walls as a house warms and cools. Glass cooking pans will crack if cooled rapidly or unevenly, because of differential contraction and the stresses it creates. (Pyrex® is less susceptible because of its small coefficient of thermal expansion.) Nuclear reactor pressure vessels are threatened by overly rapid cooling, and although none have failed, several have been cooled faster than considered desirable. Biological cells are ruptured when foods are frozen, detracting from their taste. Repeated thawing and freezing accentuate the damage. Even the oceans can be affected. A significant portion of the rise in sea level that is resulting from global warming is due to the thermal expansion of sea water.

Thermal stress contributes to the formation of potholes. (credit: Editor5807, Wikimedia Commons)

Metal is regularly used in the human body for hip and knee implants. Most implants need to be replaced over time because, among other things, metal does not bond with bone. Researchers are trying to find better metal coatings that would allow metal-to-bone bonding. One challenge is to find a coating that has an expansion coefficient similar to that of metal. If the expansion coefficients are too different, the thermal stresses during the manufacturing process lead to cracks at the coating-metal interface.

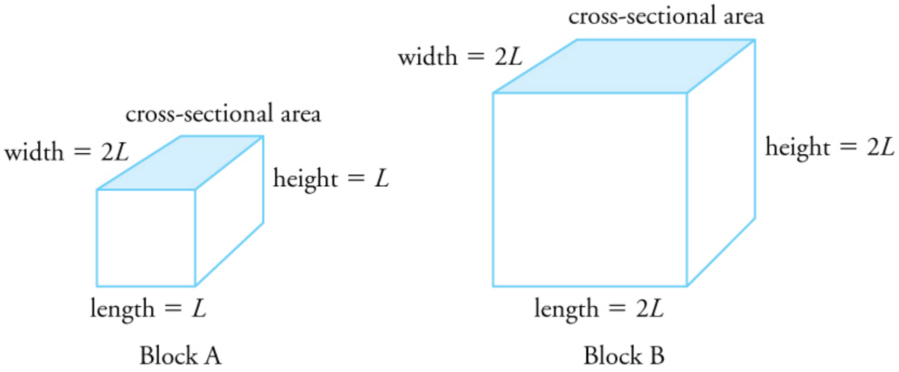
Another example of thermal stress is found in the mouth. Dental fillings can expand differently from tooth enamel. It can give pain when eating ice cream or having a hot drink. Cracks might occur in the filling. Metal fillings (gold, silver, etc.) are being replaced by composite fillings (porcelain), which have smaller coefficients of expansion, and are closer to those of teeth.

Exercise:

Check Your Understanding

Problem:

Two blocks, A and B, are made of the same material. Block A has dimensions $l \times w \times h = L \times 2L \times L$ and Block B has dimensions $2L \times 2L \times 2L$. If the temperature changes, what is (a) the change in the volume of the two blocks, (b) the change in the cross-sectional area $l \times w$, and (c) the change in the height h of the two blocks?



Solution:

(a) The change in volume is proportional to the original volume. Block A has a volume of $L \times 2L \times L = 2L^3$. Block B has a volume of $2L \times 2L \times 2L = 8L^3$, which is 4 times that of Block A. Thus the change in volume of Block B should be 4 times the change in volume of Block A.

(b) The change in area is proportional to the area. The cross-sectional area of Block A is $L \times 2L = 2L^2$, while that of Block B is

$2L \times 2L = 4L^2$. Because cross-sectional area of Block B is twice that of Block A, the change in the cross-sectional area of Block B is twice that of Block A.

(c) The change in height is proportional to the original height. Because the original height of Block B is twice that of A, the change in the height of Block B is twice that of Block A.

Section Summary

- Thermal expansion is the increase, or decrease, of the size (length, area, or volume) of a body due to a change in temperature.
- Thermal expansion is large for gases, and relatively small, but not negligible, for liquids and solids.
- Linear thermal expansion is
  Equation:
  $$\Delta L = \alpha L \Delta T,$$
  where ΔL is the change in length L, ΔT is the change in temperature, and α is the coefficient of linear expansion, which varies slightly with temperature.
- The change in area due to thermal expansion is
  Equation:
  $$\Delta A = 2\alpha A \Delta T,$$
  where ΔA is the change in area.
- The change in volume due to thermal expansion is
  Equation:
  $$\Delta V = \beta V \Delta T,$$
  where β is the coefficient of volume expansion and $\beta \approx 3\alpha$. Thermal stress is created when thermal expansion is constrained.

Conceptual Questions

Exercise:

Problem:

Thermal stresses caused by uneven cooling can easily break glass cookware. Explain why Pyrex®, a glass with a small coefficient of linear expansion, is less susceptible.

Exercise:

Problem:

Water expands significantly when it freezes: a volume increase of about 9% occurs. As a result of this expansion and because of the formation and growth of crystals as water freezes, anywhere from 10% to 30% of biological cells are burst when animal or plant material is frozen. Discuss the implications of this cell damage for the prospect of preserving human bodies by freezing so that they can be thawed at some future date when it is hoped that all diseases are curable.

Exercise:

Problem:

One method of getting a tight fit, say of a metal peg in a hole in a metal block, is to manufacture the peg slightly larger than the hole. The peg is then inserted when at a different temperature than the block. Should the block be hotter or colder than the peg during insertion? Explain your answer.

Exercise:

Problem:

Does it really help to run hot water over a tight metal lid on a glass jar before trying to open it? Explain your answer.

Exercise:

Problem:

Liquids and solids expand with increasing temperature, because the kinetic energy of a body’s atoms and molecules increases. Explain why some materials *shrink* with increasing temperature.

Problems & Exercises

Exercise:

Problem:

The height of the Washington Monument is measured to be 170 m on a day when the temperature is 35.0°C. What will its height be on a day when the temperature falls to -10.0°C? Although the monument is made of limestone, assume that its thermal coefficient of expansion is the same as marble’s.

Solution:

169.98 m

Exercise:

Problem:

How much taller does the Eiffel Tower become at the end of a day when the temperature has increased by 15°C? Its original height is 321 m and you can assume it is made of steel.

Exercise:

Problem:

What is the change in length of a 3.00-cm-long column of mercury if its temperature changes from 37.0°C to 40.0°C, assuming the mercury is unconstrained?

Solution:

$5.4 \times 10^{-6}\ \text{m}$

Exercise:

Problem:

How large an expansion gap should be left between steel railroad rails if they may reach a maximum temperature 35.0^{o}C greater than when they were laid? Their original length is 10.0 m.

Exercise:

Problem:

You are looking to purchase a small piece of land in Hong Kong. The price is "only" $60,000 per square meter! The land title says the dimensions are $20\ \text{m}\ \times\ 30\ \text{m}$. By how much would the total price change if you measured the parcel with a steel tape measure on a day when the temperature was 20^{o}C above normal?

Solution:

Because the area gets smaller, the price of the land DECREASES by ~$17,000.

Exercise:

Problem:

Global warming will produce rising sea levels partly due to melting ice caps but also due to the expansion of water as average ocean temperatures rise. To get some idea of the size of this effect, calculate the change in length of a column of water 1.00 km high for a temperature increase of 1.00^{o}C. Note that this calculation is only approximate because ocean warming is not uniform with depth.

Exercise:

Problem:

Show that 60.0 L of gasoline originally at 15.0^{o}C will expand to 61.1 L when it warms to 35.0^{o}C, as claimed in [link].

Solution:
Equation:

$$\begin{aligned} V &= V_0 + \Delta V = V_0(1 + \beta\Delta T) \\ &= (60.00\text{ L})\left[1 + \left(950 \times 10^{-6}/^{\circ}\text{C}\right)(35.0^{\circ}\text{C} - 15.0^{\circ}\text{C})\right] \\ &= 61.1\text{ L} \end{aligned}$$

Exercise:

Problem:

(a) Suppose a meter stick made of steel and one made of invar (an alloy of iron and nickel) are the same length at 0°C. What is their difference in length at 22.0°C? (b) Repeat the calculation for two 30.0-m-long surveyor's tapes.

Exercise:

Problem:

(a) If a 500-mL glass beaker is filled to the brim with ethyl alcohol at a temperature of 5.00°C, how much will overflow when its temperature reaches 22.0°C? (b) How much less water would overflow under the same conditions?

Solution:

(a) 9.35 mL

(b) 7.56 mL

Exercise:

Problem:

Most automobiles have a coolant reservoir to catch radiator fluid that may overflow when the engine is hot. A radiator is made of copper and is filled to its 16.0-L capacity when at 10.0°C. What volume of radiator fluid will overflow when the radiator and fluid reach their 95.0°C operating temperature, given that the fluid's volume coefficient of expansion is $\beta = 400 \times 10^{-6}/^{\circ}\text{C}$? Note that this coefficient is approximate, because most car radiators have operating temperatures of greater than 95.0°C.

Exercise:

Problem:

A physicist makes a cup of instant coffee and notices that, as the coffee cools, its level drops 3.00 mm in the glass cup. Show that this decrease cannot be due to thermal contraction by calculating the decrease in level if the $350\ \text{cm}^3$ of coffee is in a 7.00-cm-diameter cup and decreases in temperature from 95.0°C to 45.0°C. (Most of the drop in level is actually due to escaping bubbles of air.)

Solution:

0.832 mm

Exercise:

Problem:

(a) The density of water at 0°C is very nearly $1000\ \text{kg/m}^3$ (it is actually $999.84\ \text{kg/m}^3$), whereas the density of ice at 0°C is $917\ \text{kg/m}^3$. Calculate the pressure necessary to keep ice from expanding when it freezes, neglecting the effect such a large pressure would have on the freezing temperature. (This problem gives you only an indication of how large the forces associated with freezing water might be.) (b) What are the implications of this result for biological cells that are frozen?

Exercise:

Problem:

Show that $\beta \approx 3\alpha$, by calculating the change in volume ΔV of a cube with sides of length L.

Solution:

We know how the length changes with temperature: $\Delta L = \alpha L_0 \Delta T$. Also we know that the volume of a cube is related to its length by $V = L^3$, so the final volume is then $V = V_0 + \Delta V = (L_0 + \Delta L)^3$. Substituting for ΔL gives

Equation:

$$V = (L_0 + \alpha L_0 \Delta T)^3 = L_0^3 (1 + \alpha \Delta T)^3.$$

Now, because $\alpha \Delta T$ is small, we can use the binomial expansion:

Equation:

$$V \approx L_0^3 (1 + 3\alpha \Delta \mathrm{T}) = L_0^3 + 3\alpha L_0^3 \Delta T.$$

So writing the length terms in terms of volumes gives $V = V_0 + \Delta V \approx V_0 + 3\alpha V_0 \Delta T$, and so

Equation:

$$\Delta V = \beta V_0 \Delta T \approx 3\alpha V_0 \Delta T, \text{ or } \beta \approx 3\alpha.$$

Glossary

thermal expansion

the change in size or volume of an object with change in temperature

coefficient of linear expansion

α, the change in length, per unit length, per 1°C change in temperature; a constant used in the calculation of linear expansion; the coefficient of linear expansion depends on the material and to some degree on the temperature of the material

coefficient of volume expansion
β, the change in volume, per unit volume, per 1°C change in temperature

thermal stress
stress caused by thermal expansion or contraction

The Ideal Gas Law

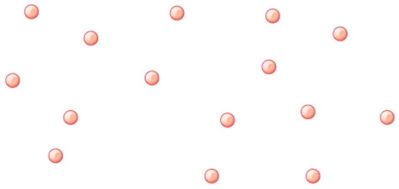
- State the ideal gas law in terms of molecules and in terms of moles.
- Use the ideal gas law to calculate pressure change, temperature change, volume change, or the number of molecules or moles in a given volume.
- Use Avogadro’s number to convert between number of molecules and number of moles.

The air inside this hot air balloon flying over Putrajaya, Malaysia, is hotter than the ambient air. As a result, the balloon experiences a buoyant force pushing it upward. (credit: Kevin Poh, Flickr)

In this section, we continue to explore the thermal behavior of gases. In particular, we examine the characteristics of atoms and molecules that compose gases. (Most gases, for example nitrogen, N_2, and oxygen, O_2, are composed of two or more atoms. We will primarily use the term “molecule” in discussing a gas because the term can also be applied to monatomic gases, such as helium.)

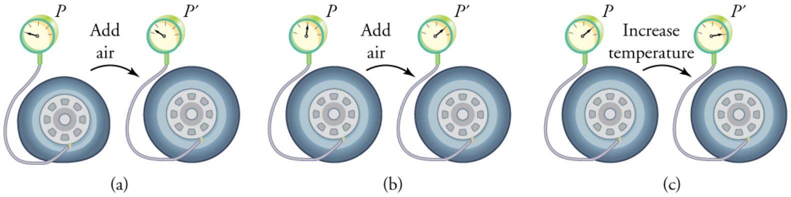
Gases are easily compressed. We can see evidence of this in [link], where you will note that gases have the *largest* coefficients of volume expansion. The large coefficients mean that gases expand and contract very rapidly with temperature changes. In addition, you will note that most gases expand at the *same* rate, or have the same β. This raises the question as to why gases should all act in nearly the same way, when liquids and solids have widely varying expansion rates.

The answer lies in the large separation of atoms and molecules in gases, compared to their sizes, as illustrated in [link]. Because atoms and molecules have large separations, forces between them can be ignored, except when they collide with each other during collisions. The motion of atoms and molecules (at temperatures well above the boiling temperature) is fast, such that the gas occupies all of the accessible volume and the expansion of gases is rapid. In contrast, in liquids and solids, atoms and molecules are closer together and are quite sensitive to the forces between them.



Atoms and molecules in a gas are typically widely separated, as shown. Because the forces between them are quite weak at these distances, the properties of a gas depend more on the number of atoms per unit volume and on temperature than on the type of atom.

To get some idea of how pressure, temperature, and volume of a gas are related to one another, consider what happens when you pump air into an initially deflated tire. The tire's volume first increases in direct proportion to the amount of air injected, without much increase in the tire pressure. Once the tire has expanded to nearly its full size, the walls limit volume expansion. If we continue to pump air into it, the pressure increases. The pressure will further increase when the car is driven and the tires move. Most manufacturers specify optimal tire pressure for cold tires. (See [link].)



(a) When air is pumped into a deflated tire, its volume first increases without much increase in pressure. (b) When the tire is filled to a certain point, the tire walls resist further expansion and the pressure increases with more air. (c) Once the tire is inflated, its pressure increases with temperature.

At room temperatures, collisions between atoms and molecules can be ignored. In this case, the gas is called an ideal gas, in which case the relationship between the pressure, volume, and temperature is given by the equation of state called the ideal gas law.

Note:

Ideal Gas Law

The **ideal gas law** states that

Equation:

$$\mathrm{PV} = \mathrm{NkT},$$

where P is the absolute pressure of a gas, V is the volume it occupies, N is the number of atoms and molecules in the gas, and T is its absolute temperature. The constant k is called the **Boltzmann constant** in honor of Austrian physicist Ludwig Boltzmann (1844–1906) and has the value

Equation:

$$k = 1.38 \times 10^{-23}\ \mathrm{J/K}.$$

The ideal gas law can be derived from basic principles, but was originally deduced from experimental measurements of Charles' law (that volume occupied by a gas is proportional to temperature at a fixed pressure) and from Boyle's law (that for a fixed temperature, the product PV is a constant). In the ideal gas model, the volume occupied by its atoms and molecules is a negligible fraction of V. The ideal gas law describes the behavior of real gases under most conditions. (Note, for example, that N is the total number of atoms and molecules, independent of the type of gas.)

Let us see how the ideal gas law is consistent with the behavior of filling the tire when it is pumped slowly and the temperature is constant. At first, the pressure P is essentially equal to atmospheric pressure, and the volume V increases in direct proportion to the number of atoms and molecules N put into the tire. Once the volume of the tire is constant, the equation $\mathrm{PV} = \mathrm{NkT}$ predicts that the pressure should increase in proportion to *the number N of atoms and molecules*.

Example:

Calculating Pressure Changes Due to Temperature Changes: Tire Pressure

Suppose your bicycle tire is fully inflated, with an absolute pressure of $7.00 \times 10^5 \text{ Pa}$ (a gauge pressure of just under 90.0 lb/in^2) at a temperature of 18.0^oC. What is the pressure after its temperature has risen to 35.0^oC? Assume that there are no appreciable leaks or changes in volume.

Strategy

The pressure in the tire is changing only because of changes in temperature. First we need to identify what we know and what we want to know, and then identify an equation to solve for the unknown.

We know the initial pressure $P_0 = 7.00 \times 10^5 \text{ Pa}$, the initial temperature $T_0 = 18.0^\text{o}\text{C}$, and the final temperature $T_\text{f} = 35.0^\text{o}\text{C}$. We must find the final pressure P_f. How can we use the equation $PV = NkT$? At first, it may seem that not enough information is given, because the volume V and number of atoms N are not specified. What we can do is use the equation twice: $P_0V_0 = NkT_0$ and $P_\text{f}V_\text{f} = NkT_\text{f}$. If we divide $P_\text{f}V_\text{f}$ by P_0V_0 we can come up with an equation that allows us to solve for P_f.

Equation:

$$\frac{P_\text{f}V_\text{f}}{P_0V_0} = \frac{N_\text{f}kT_\text{f}}{N_0kT_0}$$

Since the volume is constant, V_f and V_0 are the same and they cancel out. The same is true for N_f and N_0, and k, which is a constant. Therefore,

Equation:

$$\frac{P_\text{f}}{P_0} = \frac{T_\text{f}}{T_0}.$$

We can then rearrange this to solve for P_f:

Equation:

$$P_\text{f} = P_0\frac{T_\text{f}}{T_0},$$

where the temperature must be in units of kelvins, because T_0 and T_f are absolute temperatures.

Solution

1. Convert temperatures from Celsius to Kelvin.

Equation:

$$\begin{aligned} T_0 &= (18.0 + 273)\text{K} = 291 \text{ K} \\ T_\text{f} &= (35.0 + 273)\text{K} = 308 \text{ K} \end{aligned}$$

2. Substitute the known values into the equation.

Equation:

$$P_\text{f} = P_0\frac{T_\text{f}}{T_0} = 7.00 \times 10^5 \text{ Pa}\left(\frac{308 \text{ K}}{291 \text{ K}}\right) = 7.41 \times 10^5 \text{ Pa}$$

Discussion

The final temperature is about 6% greater than the original temperature, so the final pressure is about 6% greater as well. Note that *absolute* pressure and *absolute* temperature must be used in the ideal gas law.

Note:
Making Connections: Take-Home Experiment—Refrigerating a Balloon
Inflate a balloon at room temperature. Leave the inflated balloon in the refrigerator overnight. What happens to the balloon, and why?

Example:
Calculating the Number of Molecules in a Cubic Meter of Gas
How many molecules are in a typical object, such as gas in a tire or water in a drink? We can use the ideal gas law to give us an idea of how large N typically is.
Calculate the number of molecules in a cubic meter of gas at standard temperature and pressure (STP), which is defined to be 0ºC and atmospheric pressure.
Strategy
Because pressure, volume, and temperature are all specified, we can use the ideal gas law $\mathrm{PV}=\mathrm{NkT}$, to find N.
Solution
1. Identify the knowns.
Equation:

$$\begin{array}{lll} T & = & 0^{\circ}\mathrm{C}=273\ \mathrm{K} \\ P & = & 1.01\times 10^{5}\ \mathrm{Pa} \\ V & = & 1.00\ \mathrm{m}^{3} \\ k & = & 1.38\times 10^{-23}\ \mathrm{J/K} \end{array}$$

2. Identify the unknown: number of molecules, N.
3. Rearrange the ideal gas law to solve for N.
Equation:

$$\mathrm{PV}=\mathrm{NkT}$$
$$N=\frac{\mathrm{PV}}{\mathrm{kT}}$$

4. Substitute the known values into the equation and solve for N.
Equation:

$$N=\frac{\mathrm{PV}}{\mathrm{kT}}=\frac{\left(1.01\times 10^{5}\ \mathrm{Pa}\right)\left(1.00\ \mathrm{m}^{3}\right)}{\left(1.38\times 10^{-23}\ \mathrm{J/K}\right)\left(273\ \mathrm{K}\right)}=2.68\times 10^{25}\ \text{molecules}$$

Discussion

This number is undeniably large, considering that a gas is mostly empty space. N is huge, even in small volumes. For example, 1 cm^3 of a gas at STP has 2.68×10^{19} molecules in it. Once again, note that N is the same for all types or mixtures of gases.

Moles and Avogadro's Number

It is sometimes convenient to work with a unit other than molecules when measuring the amount of substance. A **mole** (abbreviated mol) is defined to be the amount of a substance that contains as many atoms or molecules as there are atoms in exactly 12 grams (0.012 kg) of carbon-12. The actual number of atoms or molecules in one mole is called **Avogadro's number**(N_A), in recognition of Italian scientist Amedeo Avogadro (1776–1856). He developed the concept of the mole, based on the hypothesis that equal volumes of gas, at the same pressure and temperature, contain equal numbers of molecules. That is, the number is independent of the type of gas. This hypothesis has been confirmed, and the value of Avogadro's number is

Equation:

$$N_\text{A} = 6.02 \times 10^{23}\ \text{mol}^{-1}.$$

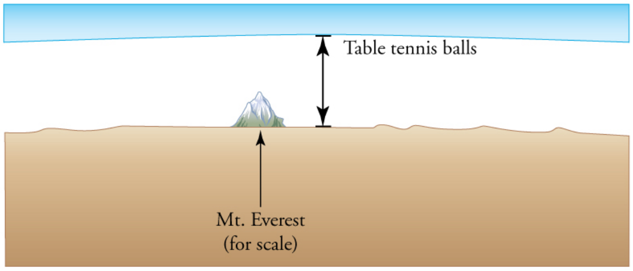
Note:

Avogadro's Number

One mole always contains 6.02×10^{23} particles (atoms or molecules), independent of the element or substance. A mole of any substance has a mass in grams equal to its molecular mass, which can be calculated from the atomic masses given in the periodic table of elements.

Equation:

$$N_\text{A} = 6.02 \times 10^{23}\ \text{mol}^{-1}$$



How big is a mole? On a macroscopic level, one mole of table tennis balls would cover the Earth to a depth of about 40 km.

Exercise:
Check Your Understanding

Problem:

The active ingredient in a Tylenol pill is 325 mg of acetaminophen $(C_8H_9NO_2)$. Find the number of active molecules of acetaminophen in a single pill.

Solution:

We first need to calculate the molar mass (the mass of one mole) of acetaminophen. To do this, we need to multiply the number of atoms of each element by the element’s atomic mass.

Equation:

$$\begin{array}{l}(8\text{ moles of carbon})(12\text{ grams/mole})+(9\text{ moles hydrogen})(1\text{ gram/mole})\\ +(1\text{ mole nitrogen})(14\text{ grams/mole})+(2\text{ moles oxygen})(16\text{ grams/mole})=151\text{ g}\end{array}$$

Then we need to calculate the number of moles in 325 mg.

Equation:

$$\left(\frac{325\text{ mg}}{151\text{ grams/mole}}\right)\left(\frac{1\text{ gram}}{1000\text{ mg}}\right)=2.15\times10^{-3}\text{ moles}$$

Then use Avogadro’s number to calculate the number of molecules.

Equation:

$$N=\left(2.15\times10^{-3}\text{ moles}\right)\left(6.02\times10^{23}\text{ molecules/mole}\right)=1.30\times10^{21}\text{ molecules}$$

Example:
Calculating Moles per Cubic Meter and Liters per Mole

Calculate: (a) the number of moles in 1.00 m^3 of gas at STP, and (b) the number of liters of gas per mole.

Strategy and Solution

(a) We are asked to find the number of moles per cubic meter, and we know from [link] that the number of molecules per cubic meter at STP is 2.68×10^{25}. The number of moles can be found by dividing the number of molecules by Avogadro’s number. We let n stand for the number of moles,

Equation:

$$n\text{ mol/m}^3=\frac{N\text{ molecules/m}^3}{6.02\times10^{23}\text{ molecules/mol}}=\frac{2.68\times10^{25}\text{ molecules/m}^3}{6.02\times10^{23}\text{ molecules/mol}}=44.5\text{ mol/m}^3.$$

(b) Using the value obtained for the number of moles in a cubic meter, and converting cubic meters to liters, we obtain

Equation:

$$\frac{\left(10^3\ \mathrm{L/m^3}\right)}{44.5\ \mathrm{mol/m^3}} = 22.5\ \mathrm{L/mol}.$$

Discussion

This value is very close to the accepted value of 22.4 L/mol. The slight difference is due to rounding errors caused by using three-digit input. Again this number is the same for all gases. In other words, it is independent of the gas.

The (average) molar weight of air (approximately 80% $\mathrm{N_2}$ and 20% $\mathrm{O_2}$ is $M = 28.8$ g. Thus the mass of one cubic meter of air is 1.28 kg. If a living room has dimensions $5\ \mathrm{m} \times 5\ \mathrm{m} \times 3\ \mathrm{m}$, the mass of air inside the room is 96 kg, which is the typical mass of a human.

Exercise:

Check Your Understanding

Problem:

The density of air at standard conditions $(P = 1\ \mathrm{atm}$ and $T = 20^\mathrm{o}\mathrm{C})$ is $1.28\ \mathrm{kg/m^3}$. At what pressure is the density $0.64\ \mathrm{kg/m^3}$ if the temperature and number of molecules are kept constant?

Solution:

The best way to approach this question is to think about what is happening. If the density drops to half its original value and no molecules are lost, then the volume must double. If we look at the equation $\mathrm{PV} = \mathrm{NkT}$, we see that when the temperature is constant, the pressure is inversely proportional to volume. Therefore, if the volume doubles, the pressure must drop to half its original value, and $P_\mathrm{f} = 0.50\ \mathrm{atm}$.

The Ideal Gas Law Restated Using Moles

A very common expression of the ideal gas law uses the number of moles, n, rather than the number of atoms and molecules, N. We start from the ideal gas law,

Equation:

$$\mathrm{PV} = \mathrm{NkT},$$

and multiply and divide the equation by Avogadro’s number N_A. This gives

Equation:

$$\mathrm{PV} = \frac{N}{N_{\mathrm{A}}} N_{\mathrm{A}} \mathrm{kT}.$$

Note that $n = N/N_{\mathrm{A}}$ is the number of moles. We define the universal gas constant $R = N_{\mathrm{A}}k$, and obtain the ideal gas law in terms of moles.

Note:
Ideal Gas Law (in terms of moles)
The ideal gas law (in terms of moles) is
Equation:

$$\mathrm{PV} = \mathrm{nRT}.$$

The numerical value of R in SI units is
Equation:

$$R = N_{\mathrm{A}}k = \left(6.02 \times 10^{23}\ \mathrm{mol}^{-1}\right)\left(1.38 \times 10^{-23}\ \mathrm{J/K}\right) = 8.31\ \mathrm{J/mol \cdot K}.$$

In other units,
Equation:

$$\begin{aligned} R &= 1.99\ \mathrm{cal/mol \cdot K} \\ R &= 0.0821\ \mathrm{L \cdot atm/mol \cdot K}. \end{aligned}$$

You can use whichever value of R is most convenient for a particular problem.

Example:
Calculating Number of Moles: Gas in a Bike Tire
How many moles of gas are in a bike tire with a volume of $2.00 \times 10^{-3}\ \mathrm{m}^3 (2.00\ \mathrm{L})$, a pressure of $7.00 \times 10^5\ \mathrm{Pa}$ (a gauge pressure of just under $90.0\ \mathrm{lb/in}^2$), and at a temperature of 18.0ºC?
Strategy
Identify the knowns and unknowns, and choose an equation to solve for the unknown. In this case, we solve the ideal gas law, $\mathrm{PV} = \mathrm{nRT}$, for the number of moles n.
Solution
1. Identify the knowns.
Equation:

$$\begin{aligned} P &= 7.00 \times 10^5\ \mathrm{Pa} \\ V &= 2.00 \times 10^{-3}\ \mathrm{m}^3 \\ T &= 18.0^{\circ}\mathrm{C} = 291\ \mathrm{K} \\ R &= 8.31\ \mathrm{J/mol \cdot K} \end{aligned}$$

2. Rearrange the equation to solve for n and substitute known values.
Equation:

$$\begin{aligned} n &= \frac{\text{PV}}{\text{RT}} = \frac{\left(7.00\times10^{5}\ \text{Pa}\right)\left(2.00\times10^{-3}\ \text{m}^{3}\right)}{(8.31\ \text{J/mol}\cdot\text{K})(291\ \text{K})} \\ &= 0.579\ \text{mol} \end{aligned}$$

Discussion
The most convenient choice for R in this case is $8.31\ \text{J/mol}\cdot\text{K}$, because our known quantities are in SI units. The pressure and temperature are obtained from the initial conditions in [link], but we would get the same answer if we used the final values.

The ideal gas law can be considered to be another manifestation of the law of conservation of energy (see Conservation of Energy). Work done on a gas results in an increase in its energy, increasing pressure and/or temperature, or decreasing volume. This increased energy can also be viewed as increased internal kinetic energy, given the gas’s atoms and molecules.

The Ideal Gas Law and Energy

Let us now examine the role of energy in the behavior of gases. When you inflate a bike tire by hand, you do work by repeatedly exerting a force through a distance. This energy goes into increasing the pressure of air inside the tire and increasing the temperature of the pump and the air.

The ideal gas law is closely related to energy: the units on both sides are joules. The right-hand side of the ideal gas law in $\text{PV}=\text{NkT}$ is NkT. This term is roughly the amount of translational kinetic energy of N atoms or molecules at an absolute temperature T, as we shall see formally in Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature. The left-hand side of the ideal gas law is PV, which also has the units of joules. We know from our study of fluids that pressure is one type of potential energy per unit volume, so pressure multiplied by volume is energy. The important point is that there is energy in a gas related to both its pressure and its volume. The energy can be changed when the gas is doing work as it expands—something we explore in Heat and Heat Transfer Methods—similar to what occurs in gasoline or steam engines and turbines.

Note:
Problem-Solving Strategy: The Ideal Gas Law
Step 1 Examine the situation to determine that an ideal gas is involved. Most gases are nearly ideal.
Step 2 Make a list of what quantities are given, or can be inferred from the problem as stated (identify the known quantities). Convert known values into proper SI units (K for temperature, Pa for pressure, m^3 for volume, molecules for N, and moles for n).
Step 3 Identify exactly what needs to be determined in the problem (identify the unknown quantities). A written list is useful.

Step 4 Determine whether the number of molecules or the number of moles is known, in order to decide which form of the ideal gas law to use. The first form is $\mathrm{PV} = \mathrm{NkT}$ and involves N, the number of atoms or molecules. The second form is $\mathrm{PV} = \mathrm{nRT}$ and involves n, the number of moles.
Step 5 Solve the ideal gas law for the quantity to be determined (the unknown quantity). You may need to take a ratio of final states to initial states to eliminate the unknown quantities that are kept fixed.
Step 6 Substitute the known quantities, along with their units, into the appropriate equation, and obtain numerical solutions complete with units. Be certain to use absolute temperature and absolute pressure.
Step 7 Check the answer to see if it is reasonable: Does it make sense?

Exercise:
Check Your Understanding

Problem:

Liquids and solids have densities about 1000 times greater than gases. Explain how this implies that the distances between atoms and molecules in gases are about 10 times greater than the size of their atoms and molecules.

Solution:

Atoms and molecules are close together in solids and liquids. In gases they are separated by empty space. Thus gases have lower densities than liquids and solids. Density is mass per unit volume, and volume is related to the size of a body (such as a sphere) cubed. So if the distance between atoms and molecules increases by a factor of 10, then the volume occupied increases by a factor of 1000, and the density decreases by a factor of 1000.

Section Summary

- The ideal gas law relates the pressure and volume of a gas to the number of gas molecules and the temperature of the gas.
- The ideal gas law can be written in terms of the number of molecules of gas:
  Equation:

  $$\mathrm{PV} = \mathrm{NkT},$$

  where P is pressure, V is volume, T is temperature, N is number of molecules, and k is the Boltzmann constant
  Equation:

  $$k = 1.38 \times 10^{-23}\ \mathrm{J/K}.$$

- A mole is the number of atoms in a 12-g sample of carbon-12.
- The number of molecules in a mole is called Avogadro’s number N_{A},

Equation:

$$N_{\mathrm{A}} = 6.02 \times 10^{23}\ \mathrm{mol}^{-1}.$$

- A mole of any substance has a mass in grams equal to its molecular weight, which can be determined from the periodic table of elements.
- The ideal gas law can also be written and solved in terms of the number of moles of gas:
  Equation:

  $$\mathrm{PV} = \mathrm{nRT},$$

  where n is number of moles and R is the universal gas constant,
  Equation:

  $$R = 8.31\ \mathrm{J/mol} \cdot \mathrm{K}.$$

- The ideal gas law is generally valid at temperatures well above the boiling temperature.

Conceptual Questions

Exercise:

Problem:

Find out the human population of Earth. Is there a mole of people inhabiting Earth? If the average mass of a person is 60 kg, calculate the mass of a mole of people. How does the mass of a mole of people compare with the mass of Earth?

Exercise:

Problem:

Under what circumstances would you expect a gas to behave significantly differently than predicted by the ideal gas law?

Exercise:

Problem:

A constant-volume gas thermometer contains a fixed amount of gas. What property of the gas is measured to indicate its temperature?

Problems & Exercises

Exercise:

Problem:

The gauge pressure in your car tires is $2.50 \times 10^5\ \mathrm{N/m^2}$ at a temperature of $35.0^{\mathrm{o}}\mathrm{C}$ when you drive it onto a ferry boat to Alaska. What is their gauge pressure later, when their temperature has dropped to $-40.0^{\mathrm{o}}\mathrm{C}$?

Solution:

1.62 atm

Exercise:

Problem:

Convert an absolute pressure of $7.00 \times 10^5\ \mathrm{N/m^2}$ to gauge pressure in $\mathrm{lb/in^2}$. (This value was stated to be just less than $90.0\ \mathrm{lb/in^2}$ in [link]. Is it?)

Exercise:

Problem:

Suppose a gas-filled incandescent light bulb is manufactured so that the gas inside the bulb is at atmospheric pressure when the bulb has a temperature of $20.0^\mathrm{o}\mathrm{C}$. (a) Find the gauge pressure inside such a bulb when it is hot, assuming its average temperature is $60.0^\mathrm{o}\mathrm{C}$ (an approximation) and neglecting any change in volume due to thermal expansion or gas leaks. (b) The actual final pressure for the light bulb will be less than calculated in part (a) because the glass bulb will expand. What will the actual final pressure be, taking this into account? Is this a negligible difference?

Solution:

(a) 0.136 atm

(b) 0.135 atm. The difference between this value and the value from part (a) is negligible.

Exercise:

Problem:

Large helium-filled balloons are used to lift scientific equipment to high altitudes. (a) What is the pressure inside such a balloon if it starts out at sea level with a temperature of $10.0^\mathrm{o}\mathrm{C}$ and rises to an altitude where its volume is twenty times the original volume and its temperature is $-50.0^\mathrm{o}\mathrm{C}$? (b) What is the gauge pressure? (Assume atmospheric pressure is constant.)

Exercise:

Problem:

Confirm that the units of nRT are those of energy for each value of R: (a) $8.31\ \mathrm{J/mol \cdot K}$, (b) $1.99\ \mathrm{cal/mol \cdot K}$, and (c) $0.0821\ \mathrm{L \cdot atm/mol \cdot K}$.

Solution:

(a) $\mathrm{nRT} = (\mathrm{mol})(\mathrm{J/mol \cdot K})(\mathrm{K}) = \mathrm{J}$

(b) $\mathrm{nRT} = (\mathrm{mol})(\mathrm{cal/mol \cdot K})(\mathrm{K}) = \mathrm{cal}$

$$\begin{array}{rcl} \mathrm{nRT} & = & (\mathrm{mol})(\mathrm{L}\cdot\mathrm{atm}/\mathrm{mol}\cdot\mathrm{K})(\mathrm{K}) \\ \text{(c)} \qquad & = & \mathrm{L}\cdot\mathrm{atm} = (\mathrm{m}^3)(\mathrm{N}/\mathrm{m}^2) \\ & = & \mathrm{N}\cdot\mathrm{m} = \mathrm{J} \end{array}$$

Exercise:

Problem:

In the text, it was shown that $N/V = 2.68 \times 10^{25}\ \mathrm{m}^{-3}$ for gas at STP. (a) Show that this quantity is equivalent to $N/V = 2.68 \times 10^{19}\ \mathrm{cm}^{-3}$, as stated. (b) About how many atoms are there in one $\mu\mathrm{m}^3$ (a cubic micrometer) at STP? (c) What does your answer to part (b) imply about the separation of atoms and molecules?

Exercise:

Problem:

Calculate the number of moles in the 2.00-L volume of air in the lungs of the average person. Note that the air is at $37.0^{\mathrm{o}}\mathrm{C}$ (body temperature).

Solution:

$7.86 \times 10^{-2}\ \mathrm{mol}$

Exercise:

Problem:

An airplane passenger has $100\ \mathrm{cm}^3$ of air in his stomach just before the plane takes off from a sea-level airport. What volume will the air have at cruising altitude if cabin pressure drops to $7.50 \times 10^4\ \mathrm{N/m}^2$?

Exercise:

Problem:

(a) What is the volume (in km^3) of Avogadro’s number of sand grains if each grain is a cube and has sides that are 1.0 mm long? (b) How many kilometers of beaches in length would this cover if the beach averages 100 m in width and 10.0 m in depth? Neglect air spaces between grains.

Solution:

(a) $6.02 \times 10^5\ \mathrm{km}^3$

(b) $6.02 \times 10^8\ \mathrm{km}$

Exercise:

Problem:

An expensive vacuum system can achieve a pressure as low as $1.00 \times 10^{-7}\ \text{N/m}^2$ at 20^{o}C. How many atoms are there in a cubic centimeter at this pressure and temperature?

Exercise:

Problem:

The number density of gas atoms at a certain location in the space above our planet is about $1.00 \times 10^{11}\ \text{m}^{-3}$, and the pressure is $2.75 \times 10^{-10}\ \text{N/m}^2$ in this space. What is the temperature there?

Solution:

-73.9^{o}C

Exercise:

Problem:

A bicycle tire has a pressure of $7.00 \times 10^{5}\ \text{N/m}^2$ at a temperature of 18.0^{o}C and contains 2.00 L of gas. What will its pressure be if you let out an amount of air that has a volume of $100\ \text{cm}^3$ at atmospheric pressure? Assume tire temperature and volume remain constant.

Exercise:

Problem:

A high-pressure gas cylinder contains 50.0 L of toxic gas at a pressure of $1.40 \times 10^{7}\ \text{N/m}^2$ and a temperature of 25.0^{o}C. Its valve leaks after the cylinder is dropped. The cylinder is cooled to dry ice temperature $(–78.5^{\text{o}}\text{C})$ to reduce the leak rate and pressure so that it can be safely repaired. (a) What is the final pressure in the tank, assuming a negligible amount of gas leaks while being cooled and that there is no phase change? (b) What is the final pressure if one-tenth of the gas escapes? (c) To what temperature must the tank be cooled to reduce the pressure to 1.00 atm (assuming the gas does not change phase and that there is no leakage during cooling)? (d) Does cooling the tank appear to be a practical solution?

Solution:

(a) $9.14 \times 10^{6}\ \text{N/m}^2$

(b) $8.23 \times 10^{6}\ \text{N/m}^2$

(c) 2.16 K

(d) No. The final temperature needed is much too low to be easily achieved for a large object.

Exercise:

Problem:

Find the number of moles in 2.00 L of gas at $35.0^{\mathrm{o}}\mathrm{C}$ and under $7.41 \times 10^{7}\ \mathrm{N/m^{2}}$ of pressure.

Exercise:

Problem:

Calculate the depth to which Avogadro's number of table tennis balls would cover Earth. Each ball has a diameter of 3.75 cm. Assume the space between balls adds an extra 25.0% to their volume and assume they are not crushed by their own weight.

Solution:

41 km

Exercise:

Problem:

(a) What is the gauge pressure in a $25.0^{\mathrm{o}}\mathrm{C}$ car tire containing 3.60 mol of gas in a 30.0 L volume? (b) What will its gauge pressure be if you add 1.00 L of gas originally at atmospheric pressure and $25.0^{\mathrm{o}}\mathrm{C}$? Assume the temperature returns to $25.0^{\mathrm{o}}\mathrm{C}$ and the volume remains constant.

Exercise:

Problem:

(a) In the deep space between galaxies, the density of atoms is as low as $10^{6}\ \mathrm{atoms/m^{3}}$, and the temperature is a frigid 2.7 K. What is the pressure? (b) What volume (in $\mathrm{m^{3}}$) is occupied by 1 mol of gas? (c) If this volume is a cube, what is the length of its sides in kilometers?

Solution:

(a) $3.7 \times 10^{-17}\ \mathrm{Pa}$

(b) $6.0 \times 10^{17}\ \mathrm{m^{3}}$

(c) $8.4 \times 10^{2}\ \mathrm{km}$

Glossary

ideal gas law
: the physical law that relates the pressure and volume of a gas to the number of gas molecules or number of moles of gas and the temperature of the gas

Boltzmann constant

k , a physical constant that relates energy to temperature; $k = 1.38 \times 10^{-23}\ \text{J/K}$

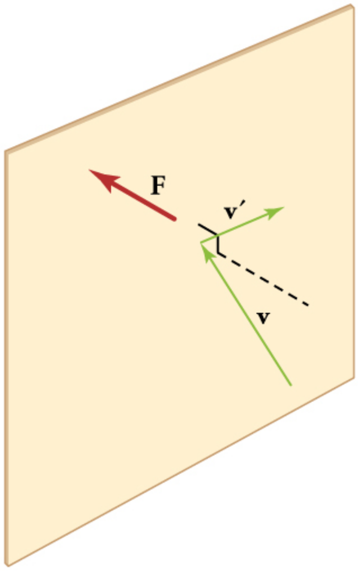
Avogadro’s number
N_A , the number of molecules or atoms in one mole of a substance; $N_\text{A} = 6.02 \times 10^{23}$ particles/mole

mole
the quantity of a substance whose mass (in grams) is equal to its molecular mass

Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature

- Express the ideal gas law in terms of molecular mass and velocity.
- Define thermal energy.
- Calculate the kinetic energy of a gas molecule, given its temperature.
- Describe the relationship between the temperature of a gas and the kinetic energy of atoms and molecules.
- Describe the distribution of speeds of molecules in a gas.

We have developed macroscopic definitions of pressure and temperature. Pressure is the force divided by the area on which the force is exerted, and temperature is measured with a thermometer. We gain a better understanding of pressure and temperature from the kinetic theory of gases, which assumes that atoms and molecules are in continuous random motion.



When a molecule collides with a rigid wall, the component of its momentum perpendicular to the wall is reversed. A force is thus exerted on the wall, creating pressure.

[link] shows an elastic collision of a gas molecule with the wall of a container, so that it exerts a force on the wall (by Newton's third law). Because a huge number of molecules will collide with the wall in a short time, we observe an average force per unit area. These collisions are the source of pressure in a gas. As the number of molecules increases, the number of collisions and thus the pressure increase. Similarly, the gas pressure is higher if the average velocity of molecules is higher. The actual relationship is derived in the Things Great and Small feature below. The following relationship is found:

Equation:

$$\text{PV} = \frac{1}{3}\text{Nm}\overline{v^2},$$

where P is the pressure (average force per unit area), V is the volume of gas in the container, N is the number of molecules in the container, m is the mass of a molecule, and $\overline{v^2}$ is the average of the molecular speed squared.

What can we learn from this atomic and molecular version of the ideal gas law? We can derive a relationship between temperature and the average translational kinetic energy of molecules in a gas. Recall the previous expression of the ideal gas law:

Equation:

$$\text{PV} = \text{NkT}.$$

Equating the right-hand side of this equation with the right-hand side of $\text{PV} = \frac{1}{3}\text{Nm}\overline{v^2}$ gives

Equation:

$$\frac{1}{3}\text{Nm}\overline{v^2} = \text{NkT}.$$

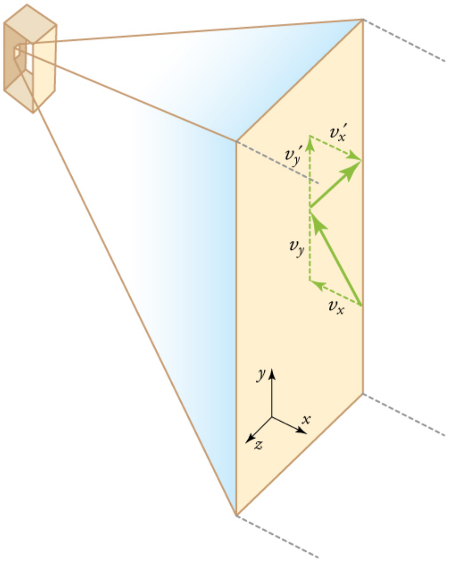
Note:

Making Connections: Things Great and Small—Atomic and Molecular Origin of Pressure in a Gas

[link] shows a box filled with a gas. We know from our previous discussions that putting more gas into the box produces greater pressure, and that increasing the temperature of the gas also produces a greater pressure. But why should increasing the temperature of the gas increase the pressure in the box? A look at the atomic and

molecular scale gives us some answers, and an alternative expression for the ideal gas law.

The figure shows an expanded view of an elastic collision of a gas molecule with the wall of a container. Calculating the average force exerted by such molecules will lead us to the ideal gas law, and to the connection between temperature and molecular kinetic energy. We assume that a molecule is small compared with the separation of molecules in the gas, and that its interaction with other molecules can be ignored. We also assume the wall is rigid and that the molecule's direction changes, but that its speed remains constant (and hence its kinetic energy and the magnitude of its momentum remain constant as well). This assumption is not always valid, but the same result is obtained with a more detailed description of the molecule's exchange of energy and momentum with the wall.



Gas in a box exerts an outward pressure on its walls. A molecule colliding with a rigid wall has the direction of its velocity and momentum in the x-direction reversed. This direction is perpendicular to the wall. The components of its velocity momentum in the y- and z-directions are not changed, which means there is no force parallel to the wall.

If the molecule’s velocity changes in the x-direction, its momentum changes from $-\mathrm{mv}_x$ to $+\mathrm{mv}_x$. Thus, its change in momentum is $\Delta \mathrm{mv} = +\mathrm{mv}_x - (-\mathrm{mv}_x) = 2\mathrm{mv}_x$. The force exerted on the molecule is given by

Equation:

$$F = \frac{\Delta p}{\Delta t} = \frac{2\mathrm{mv}_x}{\Delta t}.$$

There is no force between the wall and the molecule until the molecule hits the wall. During the short time of the collision, the force between the molecule and wall is relatively large. We are looking for an average force; we take Δt to be the average time between collisions of the molecule with this wall. It is the time it would take the molecule to go across the box and back (a distance $2l$) at a speed of v_x. Thus $\Delta t = 2l/v_x$, and the expression for the force becomes

Equation:

$$F = \frac{2\mathrm{mv}_x}{2l/v_x} = \frac{mv_x^2}{l}.$$

This force is due to *one* molecule. We multiply by the number of molecules N and use their average squared velocity to find the force

Equation:

$$F = N\frac{m\overline{v_x^2}}{l},$$

where the bar over a quantity means its average value. We would like to have the force in terms of the speed v, rather than the x-component of the velocity. We note that the total velocity squared is the sum of the squares of its components, so that

Equation:

$$\overline{v^2} = \overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}.$$

Because the velocities are random, their average components in all directions are the same:

Equation:

$$\overline{v_x^2} = \overline{v_y^2} = \overline{v_z^2}.$$

Thus,

Equation:

$$\overline{v^2} = 3\overline{v_x^2},$$

or

Equation:

$$\overline{v_x^2} = \frac{1}{3}\overline{v^2}.$$

Substituting $\frac{1}{3}\overline{v^2}$ into the expression for F gives

Equation:

$$F = N\frac{m\overline{v^2}}{3l}.$$

The pressure is $F/A,$ so that we obtain

Equation:

$$P = \frac{F}{A} = N\frac{m\overline{v^2}}{3\text{Al}} = \frac{1}{3}\frac{\text{Nm}\overline{v^2}}{V},$$

where we used $V = \text{Al}$ for the volume. This gives the important result.

Equation:

$$\text{PV} = \frac{1}{3}\text{Nm}\overline{v^2}$$

This equation is another expression of the ideal gas law.

We can get the average kinetic energy of a molecule, $\frac{1}{2}mv^2$, from the right-hand side of the equation by canceling N and multiplying by 3/2. This calculation produces the result that the average kinetic energy of a molecule is directly related to absolute temperature.

Equation:

$$\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}\text{kT}$$

The average translational kinetic energy of a molecule, $\overline{\text{KE}}$, is called **thermal energy.** The equation $\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}\text{kT}$ is a molecular interpretation of temperature, and it has been found to be valid for gases and reasonably accurate in liquids and solids. It is another definition of temperature based on an expression of the molecular energy.

It is sometimes useful to rearrange $\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}\text{kT}$**,** and solve for the average speed of molecules in a gas in terms of temperature,
Equation:

$$\sqrt{\overline{v^2}} = v_{\text{rms}} = \sqrt{\frac{3\text{kT}}{m}},$$

where v_{rms} stands for root-mean-square (rms) speed.

Example:
Calculating Kinetic Energy and Speed of a Gas Molecule
(a) What is the average kinetic energy of a gas molecule at 20.0^{o}C (room temperature)? (b) Find the rms speed of a nitrogen molecule (N_2) at this temperature.
Strategy for (a)
The known in the equation for the average kinetic energy is the temperature.
Equation:

$$\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}\text{kT}$$

Before substituting values into this equation, we must convert the given temperature to kelvins. This conversion gives $T = (20.0 + 273)\text{ K} = 293\text{ K.}$
Solution for (a)
The temperature alone is sufficient to find the average translational kinetic energy. Substituting the temperature into the translational kinetic energy equation gives
Equation:

$$\overline{\text{KE}} = \frac{3}{2}\text{kT} = \frac{3}{2}\left(1.38 \times 10^{-23}\text{ J/K}\right)(293\text{ K}) = 6.07 \times 10^{-21}\text{ J.}$$

Strategy for (b)

Finding the rms speed of a nitrogen molecule involves a straightforward calculation using the equation

Equation:

$$\sqrt{\overline{v^2}} = v_{\text{rms}} = \sqrt{\frac{3\text{kT}}{m}},$$

but we must first find the mass of a nitrogen molecule. Using the molecular mass of nitrogen N_2 from the periodic table,

Equation:

$$m = \frac{2(14.0067) \times 10^{-3}\ \text{kg/mol}}{6.02 \times 10^{23}\ \text{mol}^{-1}} = 4.65 \times 10^{-26}\ \text{kg}.$$

Solution for (b)

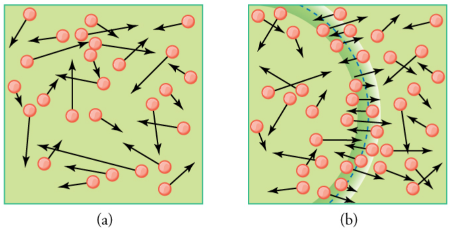
Substituting this mass and the value for k into the equation for v_{rms} yields

Equation:

$$v_{\text{rms}} = \sqrt{\frac{3\text{kT}}{m}} = \sqrt{\frac{3\left(1.38 \times 10^{-23}\ \text{J/K}\right)(293\ \text{K})}{4.65 \times 10^{-26}\ \text{kg}}} = 511\ \text{m/s}.$$

Discussion

Note that the average kinetic energy of the molecule is independent of the type of molecule. The average translational kinetic energy depends only on absolute temperature. The kinetic energy is very small compared to macroscopic energies, so that we do not feel when an air molecule is hitting our skin. The rms velocity of the nitrogen molecule is surprisingly large. These large molecular velocities do not yield macroscopic movement of air, since the molecules move in all directions with equal likelihood. The *mean free path* (the distance a molecule can move on average between collisions) of molecules in air is very small, and so the molecules move rapidly but do not get very far in a second. The high value for rms speed is reflected in the speed of sound, however, which is about 340 m/s at room temperature. The faster the rms speed of air molecules, the faster that sound vibrations can be transferred through the air. The speed of sound increases with temperature and is greater in gases with small molecular masses, such as helium. (See [link].)

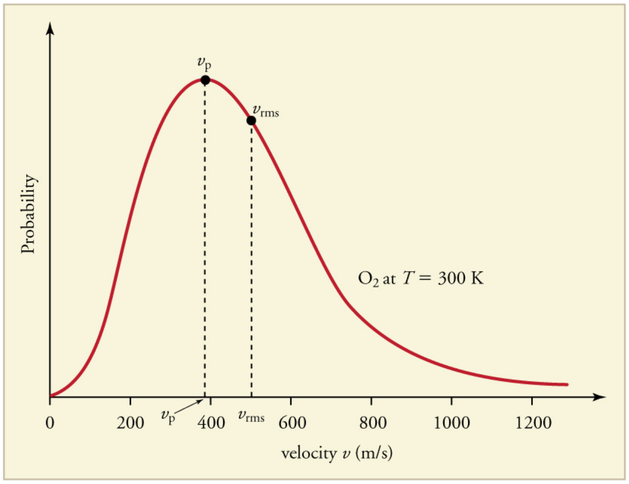


(a) There are many molecules moving so fast in an ordinary gas that they collide a billion times every second. (b) Individual molecules do not move very far in a small amount of time, but disturbances like sound waves are transmitted at speeds related to the molecular speeds.

Note:
Making Connections: Historical Note—Kinetic Theory of Gases
The kinetic theory of gases was developed by Daniel Bernoulli (1700–1782), who is best known in physics for his work on fluid flow (hydrodynamics). Bernoulli's work predates the atomistic view of matter established by Dalton.

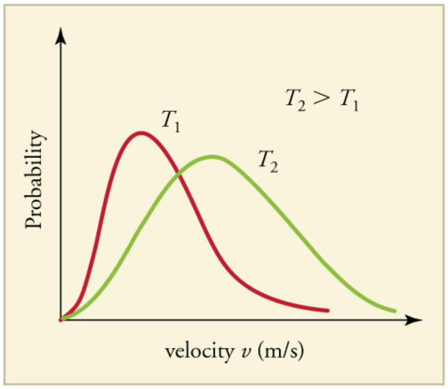
Distribution of Molecular Speeds

The motion of molecules in a gas is random in magnitude and direction for individual molecules, but a gas of many molecules has a predictable distribution of molecular speeds. This distribution is called the *Maxwell-Boltzmann distribution*, after its originators, who calculated it based on kinetic theory, and has since been confirmed experimentally. (See [link].) The distribution has a long tail, because a few molecules may go several times the rms speed. The most probable speed v_{p} is less than the rms speed v_{rms}. [link] shows that the curve is shifted to higher speeds at higher temperatures, with a broader range of speeds.



The Maxwell-Boltzmann distribution of molecular speeds in an ideal gas. The most likely speed v_p is less than the rms speed v_{rms}. Although very high speeds are possible, only a tiny fraction of the molecules have speeds that are an order of magnitude greater than v_{rms}.

The distribution of thermal speeds depends strongly on temperature. As temperature increases, the speeds are shifted to higher values and the distribution is broadened.



The Maxwell-Boltzmann distribution is shifted to

higher speeds and is
broadened at higher
temperatures.

What is the implication of the change in distribution with temperature shown in [link] for humans? All other things being equal, if a person has a fever, he or she is likely to lose more water molecules, particularly from linings along moist cavities such as the lungs and mouth, creating a dry sensation in the mouth.

Example:

Calculating Temperature: Escape Velocity of Helium Atoms

In order to escape Earth’s gravity, an object near the top of the atmosphere (at an altitude of 100 km) must travel away from Earth at 11.1 km/s. This speed is called the *escape velocity*. At what temperature would helium atoms have an rms speed equal to the escape velocity?

Strategy

Identify the knowns and unknowns and determine which equations to use to solve the problem.

Solution

1. Identify the knowns: v is the escape velocity, 11.1 km/s.
2. Identify the unknowns: We need to solve for temperature, T. We also need to solve for the mass m of the helium atom.
3. Determine which equations are needed.

- To solve for mass m of the helium atom, we can use information from the periodic table:
  Equation:

$$m=\frac{\text{molar mass}}{\text{number of atoms per mole}}.$$

- To solve for temperature T, we can rearrange either
  Equation:

$$\overline{\text{KE}}=\frac{1}{2}m\overline{v^2}=\frac{3}{2}\text{kT}$$

or

Equation:

$$\sqrt{\overline{v^2}} = v_{\text{rms}} = \sqrt{\frac{3\text{kT}}{m}}$$

to yield

Equation:

$$T = \frac{m\overline{v^2}}{3k},$$

where k is the Boltzmann constant and m is the mass of a helium atom.

4. Plug the known values into the equations and solve for the unknowns.

Equation:

$$m = \frac{\text{molar mass}}{\text{number of atoms per mole}} = \frac{4.0026 \times 10^{-3}\ \text{kg/mol}}{6.02 \times 10^{23}\ \text{mol}} = 6.65 \times 10^{-27}\ \text{kg}$$

Equation:

$$T = \frac{\left(6.65 \times 10^{-27}\ \text{kg}\right)\left(11.1 \times 10^{3}\ \text{m/s}\right)^2}{3\left(1.38 \times 10^{-23}\ \text{J/K}\right)} = 1.98 \times 10^{4}\ \text{K}$$

Discussion

This temperature is much higher than atmospheric temperature, which is approximately 250 K $\left(-25^{\text{o}}\text{C or } -10^{\text{o}}\text{F}\right)$ at high altitude. Very few helium atoms are left in the atmosphere, but there were many when the atmosphere was formed. The reason for the loss of helium atoms is that there are a small number of helium atoms with speeds higher than Earth’s escape velocity even at normal temperatures. The speed of a helium atom changes from one instant to the next, so that at any instant, there is a small, but nonzero chance that the speed is greater than the escape speed and the molecule escapes from Earth’s gravitational pull. Heavier molecules, such as oxygen, nitrogen, and water (very little of which reach a very high altitude), have smaller rms speeds, and so it is much less likely that any of them will have speeds greater than the escape velocity. In fact, so few have speeds above the escape velocity that billions of years are required to lose significant amounts of the atmosphere. [link] shows the impact of a lack of an atmosphere on the Moon. Because the gravitational pull of the Moon is much weaker, it has lost almost its entire atmosphere. The comparison between Earth and the Moon is discussed in this chapter’s Problems and Exercises.

This photograph of Apollo 17 Commander Eugene Cernan driving the lunar rover on the Moon in 1972 looks as though it was taken at night with a large spotlight. In fact, the light is coming from the Sun. Because the acceleration due to gravity on the Moon is so low (about 1/6 that of Earth), the Moon’s escape velocity is much smaller. As a result, gas molecules escape very easily from the Moon, leaving it with virtually no atmosphere. Even during the daytime, the sky is black because there is no gas to scatter sunlight. (credit: Harrison H. Schmitt/NASA)

Exercise:
Check Your Understanding

Problem:

If you consider a very small object such as a grain of pollen, in a gas, then the number of atoms and molecules striking its surface would also be relatively small. Would the grain of pollen experience any fluctuations in pressure due to statistical fluctuations in the number of gas atoms and molecules striking it in a given amount of time?

Solution:

Yes. Such fluctuations actually occur for a body of any size in a gas, but since the numbers of atoms and molecules are immense for macroscopic bodies, the fluctuations are a tiny percentage of the number of collisions, and the averages spoken of in this section vary imperceptibly. Roughly speaking the fluctuations are proportional to the inverse square root of the number of collisions, so for small bodies they can become significant. This was actually observed in the 19th century for pollen grains in water, and is known as the Brownian effect.

Note:
PhET Explorations: Gas Properties
Pump gas molecules into a box and see what happens as you change the volume, add or remove heat, change gravity, and more. Measure the temperature and pressure, and discover how the properties of the gas vary in relation to each other.

Gas Properties

Section Summary

- Kinetic theory is the atomistic description of gases as well as liquids and solids.

- Kinetic theory models the properties of matter in terms of continuous random motion of atoms and molecules.
- The ideal gas law can also be expressed as
  Equation:

  $$\text{PV} = \frac{1}{3}\text{Nm}\overline{v^2},$$

  where P is the pressure (average force per unit area), V is the volume of gas in the container, N is the number of molecules in the container, m is the mass of a molecule, and $\overline{v^2}$ is the average of the molecular speed squared.
- Thermal energy is defined to be the average translational kinetic energy $\overline{\text{KE}}$ of an atom or molecule.
- The temperature of gases is proportional to the average translational kinetic energy of atoms and molecules.
  Equation:

  $$\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}\text{kT}$$

  or
  Equation:

  $$\sqrt{\overline{v^2}} = v_{\text{rms}} = \sqrt{\frac{3\text{kT}}{m}}.$$
- The motion of individual molecules in a gas is random in magnitude and direction. However, a gas of many molecules has a predictable distribution of molecular speeds, known as the *Maxwell-Boltzmann distribution*.

Conceptual Questions

Exercise:

Problem:

How is momentum related to the pressure exerted by a gas? Explain on the atomic and molecular level, considering the behavior of atoms and molecules.

Problems & Exercises

Exercise:

Problem:

Some incandescent light bulbs are filled with argon gas. What is v_{rms} for argon atoms near the filament, assuming their temperature is 2500 K?

Solution:

$1.25\times 10^{3}\ \text{m/s}$

Exercise:

Problem:

Average atomic and molecular speeds (v_{rms}) are large, even at low temperatures. What is v_{rms} for helium atoms at 5.00 K, just one degree above helium's liquefaction temperature?

Exercise:

Problem:

(a) What is the average kinetic energy in joules of hydrogen atoms on the 5500^{o}C surface of the Sun? (b) What is the average kinetic energy of helium atoms in a region of the solar corona where the temperature is $6.00\times 10^{5}\ \text{K}$?

Solution:

(a) $1.20\times 10^{-19}\ \text{J}$

(b) $1.24\times 10^{-17}\ \text{J}$

Exercise:

Problem:

The escape velocity of any object from Earth is 11.2 km/s. (a) Express this speed in m/s and km/h. (b) At what temperature would oxygen molecules (molecular mass is equal to 32.0 g/mol) have an average velocity v_{rms} equal to Earth's escape velocity of 11.1 km/s?

Exercise:

Problem:

The escape velocity from the Moon is much smaller than from Earth and is only 2.38 km/s. At what temperature would hydrogen molecules (molecular mass is equal to 2.016 g/mol) have an average velocity v_{rms} equal to the Moon's escape velocity?

Solution:

458 K

Exercise:

Problem:

Nuclear fusion, the energy source of the Sun, hydrogen bombs, and fusion reactors, occurs much more readily when the average kinetic energy of the atoms is high—that is, at high temperatures. Suppose you want the atoms in your fusion experiment to have average kinetic energies of $6.40 \times 10^{–14}\text{ J}$. What temperature is needed?

Exercise:

Problem:

Suppose that the average velocity (v_{rms}) of carbon dioxide molecules (molecular mass is equal to 44.0 g/mol) in a flame is found to be $1.05 \times 10^5\text{ m/s}$. What temperature does this represent?

Solution:

$1.95 \times 10^7\text{ K}$

Exercise:

Problem:

Hydrogen molecules (molecular mass is equal to 2.016 g/mol) have an average velocity v_{rms} equal to 193 m/s. What is the temperature?

Exercise:

Problem:

Much of the gas near the Sun is atomic hydrogen. Its temperature would have to be $1.5 \times 10^7 \text{ K}$ for the average velocity v_{rms} to equal the escape velocity from the Sun. What is that velocity?

Solution:

$6.09 \times 10^5 \text{ m/s}$

Exercise:

Problem:

There are two important isotopes of uranium— ^{235}U and ^{238}U; these isotopes are nearly identical chemically but have different atomic masses. Only ^{235}U is very useful in nuclear reactors. One of the techniques for separating them (gas diffusion) is based on the different average velocities v_{rms} of uranium hexafluoride gas, UF_6. (a) The molecular masses for $^{235}\text{U UF}_6$ and $^{238}\text{U UF}_6$ are 349.0 g/mol and 352.0 g/mol, respectively. What is the ratio of their average velocities? (b) At what temperature would their average velocities differ by 1.00 m/s? (c) Do your answers in this problem imply that this technique may be difficult?

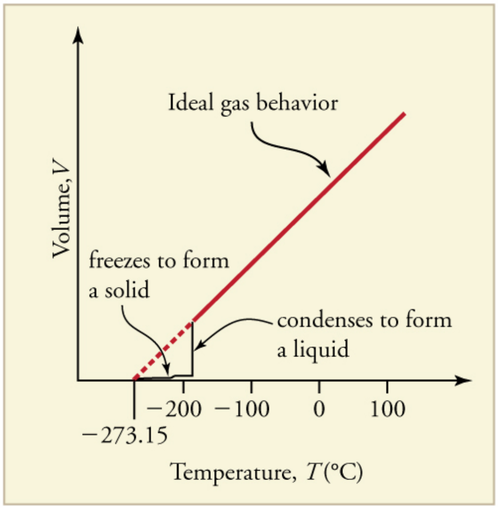
Glossary

thermal energy	$\overline{\text{KE}}$, the average translational kinetic energy of a molecule

Phase Changes

- Interpret a phase diagram.
- State Dalton's law.
- Identify and describe the triple point of a gas from its phase diagram.
- Describe the state of equilibrium between a liquid and a gas, a liquid and a solid, and a gas and a solid.

Up to now, we have considered the behavior of ideal gases. Real gases are like ideal gases at high temperatures. At lower temperatures, however, the interactions between the molecules and their volumes cannot be ignored. The molecules are very close (condensation occurs) and there is a dramatic decrease in volume, as seen in [link]. The substance changes from a gas to a liquid. When a liquid is cooled to even lower temperatures, it becomes a solid. The volume never reaches zero because of the finite volume of the molecules.



A sketch of volume versus temperature for a real gas at constant pressure. The linear (straight line) part of the graph represents ideal gas behavior—volume and temperature are directly and positively related and the line extrapolates to zero volume at -273.15°C, or absolute zero. When the gas becomes a liquid, however, the volume actually decreases precipitously at the liquefaction point. The volume decreases slightly once the substance is solid, but it never becomes zero.

High pressure may also cause a gas to change phase to a liquid. Carbon dioxide, for example, is a gas at room temperature and atmospheric pressure, but becomes a liquid under sufficiently high pressure. If the pressure is reduced, the temperature drops and the liquid carbon dioxide solidifies into a snow-like substance at the temperature -78°C. Solid CO_2 is called “dry ice.” Another example of a gas that can be in a liquid phase is liquid nitrogen (LN_2). LN_2 is made by liquefaction of atmospheric air (through compression and cooling). It boils at 77 K (-196°C) at atmospheric pressure. LN_2 is useful as a refrigerant and allows for the preservation of blood, sperm, and other biological materials. It is also used to reduce noise in electronic sensors and equipment, and to help cool down their current-carrying wires. In dermatology, LN_2 is used to freeze and painlessly remove warts and other growths from the skin.

PV Diagrams

We can examine aspects of the behavior of a substance by plotting a graph of pressure versus volume, called a ***PV* diagram**. When the substance behaves like an ideal gas, the ideal gas law describes the relationship between its pressure and volume. That is,

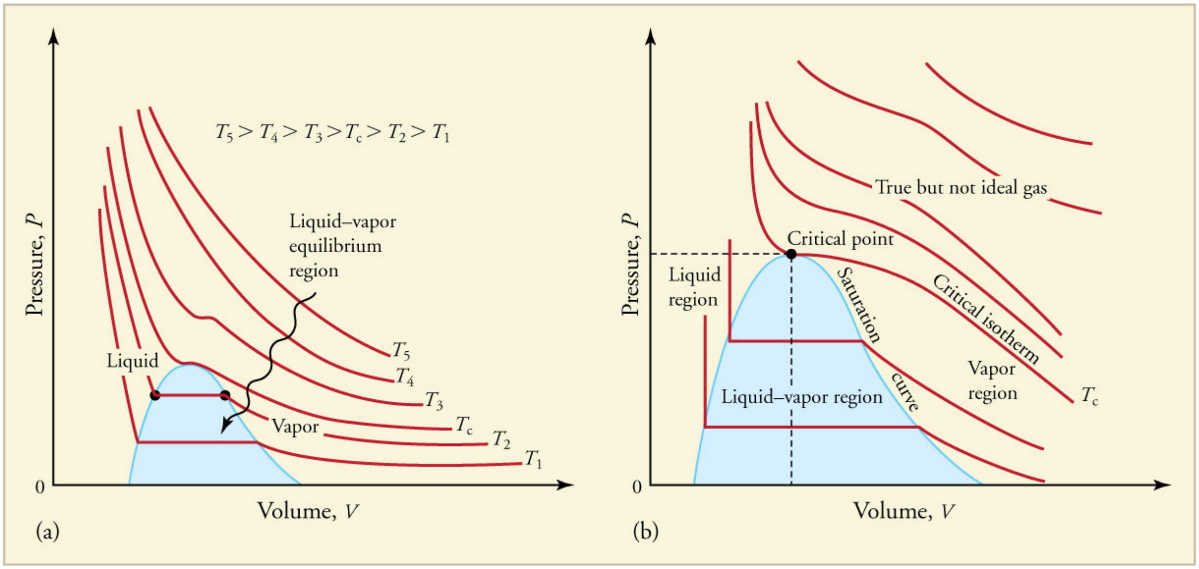
Equation:

$$\text{PV} = \text{NkT (ideal gas).}$$

Now, assuming the number of molecules and the temperature are fixed,
Equation:

$$\text{PV} = \text{constant (ideal gas, constant temperature).}$$

For example, the volume of the gas will decrease as the pressure increases. If you plot the relationship $\text{PV} = \text{constant}$ on a PV diagram, you find a hyperbola. [link] shows a graph of pressure versus volume. The hyperbolas represent ideal-gas behavior at various fixed temperatures, and are called *isotherms*. At lower temperatures, the curves begin to look less like hyperbolas—the gas is not behaving ideally and may even contain liquid. There is a **critical point**—that is, a **critical temperature**—above which liquid cannot exist. At sufficiently high pressure above the critical point, the gas will have the density of a liquid but will not condense. Carbon dioxide, for example, cannot be liquefied at a temperature above 31.0°C. **Critical pressure** is the minimum pressure needed for liquid to exist at the critical temperature. [link] lists representative critical temperatures and pressures.



PV diagrams. (a) Each curve (isotherm) represents the relationship between P and V at a fixed temperature; the upper curves are at higher temperatures. The lower curves are not hyperbolas, because the gas is no longer an ideal gas. (b) An expanded portion of the PV diagram for low temperatures, where the phase can change from a gas to a liquid. The term “vapor” refers to the gas phase when it exists at a temperature below the boiling temperature.

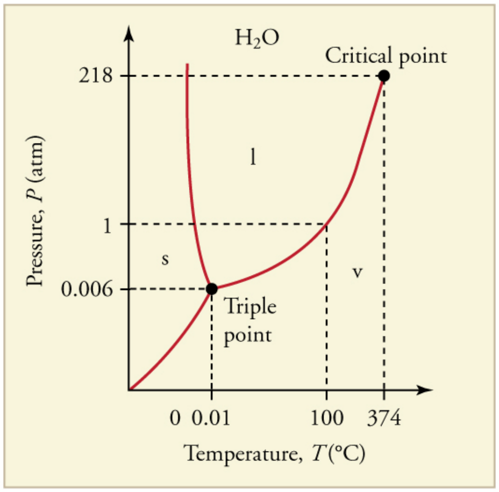
Substance	**Critical temperature**		**Critical pressure**	
	K	°C	Pa	atm
Water	647.4	374.3	22.12×10^6	219.0
Sulfur dioxide	430.7	157.6	7.88×10^6	78.0
Ammonia	405.5	132.4	11.28×10^6	111.7
Carbon dioxide	304.2	31.1	7.39×10^6	73.2

Substance	**Critical temperature**		**Critical pressure**	
	K	$^\circ\text{C}$	Pa	atm
Oxygen	154.8	−118.4	5.08×10^6	50.3
Nitrogen	126.2	−146.9	3.39×10^6	33.6
Hydrogen	33.3	−239.9	1.30×10^6	12.9
Helium	5.3	−267.9	0.229×10^6	2.27

Critical Temperatures and Pressures

Phase Diagrams

The plots of pressure versus temperatures provide considerable insight into thermal properties of substances. There are well-defined regions on these graphs that correspond to various phases of matter, so PT graphs are called **phase diagrams**. [link] shows the phase diagram for water. Using the graph, if you know the pressure and temperature you can determine the phase of water. The solid lines—boundaries between phases—indicate temperatures and pressures at which the phases coexist (that is, they exist together in ratios, depending on pressure and temperature). For example, the boiling point of water is 100°C at 1.00 atm. As the pressure increases, the boiling temperature rises steadily to 374°C at a pressure of 218 atm. A pressure cooker (or even a covered pot) will cook food faster because the water can exist as a liquid at temperatures greater than 100°C without all boiling away. The curve ends at a point called the *critical point*, because at higher temperatures the liquid phase does not exist at any pressure. The critical point occurs at the critical temperature, as you can see for water from [link]. The critical temperature for oxygen is -118°C, so oxygen cannot be liquefied above this temperature.



The phase diagram (PT graph) for water. Note that the axes are nonlinear and the graph is not to scale. This graph is simplified—there are several other exotic phases of ice at higher pressures.

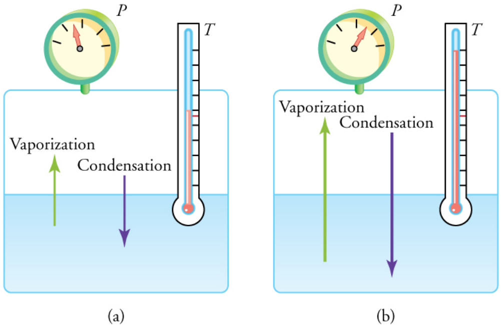
Similarly, the curve between the solid and liquid regions in [link] gives the melting temperature at various pressures. For example, the melting point is 0°C at 1.00 atm, as expected. Note that, at a fixed temperature, you can change the phase from solid (ice) to liquid (water) by increasing the pressure. Ice melts from pressure in the hands of a snowball maker. From the phase diagram, we can also say that the melting temperature of ice rises with increased pressure. When a car is driven over snow, the increased pressure from the tires melts the snowflakes; afterwards the water refreezes and forms an ice layer.

At sufficiently low pressures there is no liquid phase, but the substance can exist as either gas or solid. For water, there is no liquid phase at pressures below 0.00600 atm. The phase change from solid to gas is called **sublimation**. It accounts for large losses of snow pack that never make it into a river, the routine automatic defrosting of a freezer, and the freeze-drying process applied to many foods. Carbon dioxide, on the other hand, sublimates at standard atmospheric pressure of 1 atm. (The solid form of CO_2 is known as dry ice because it does not melt. Instead, it moves directly from the solid to the gas state.)

All three curves on the phase diagram meet at a single point, the **triple point**, where all three phases exist in equilibrium. For water, the triple point occurs at $273.16\text{ K}\ (0.01^\circ\text{C})$, and is a more accurate calibration temperature than the melting point of water at 1.00 atm, or $273.15\text{ K}\ (0.0^\circ\text{C})$. See [link] for the triple point values of other substances.

Equilibrium

Liquid and gas phases are in equilibrium at the boiling temperature. (See [link].) If a substance is in a closed container at the boiling point, then the liquid is boiling and the gas is condensing at the same rate without net change in their relative amount. Molecules in the liquid escape as a gas at the same rate at which gas molecules stick to the liquid, or form droplets and become part of the liquid phase. The combination of temperature and pressure has to be “just right”; if the temperature and pressure are increased, equilibrium is maintained by the same increase of boiling and condensation rates.



Equilibrium between liquid and gas at two different boiling points inside a closed container. (a) The rates of boiling and condensation are equal at this combination of temperature and pressure, so the liquid and gas phases are in equilibrium. (b) At a higher temperature, the boiling rate is faster and the rates at which molecules leave the liquid and enter the gas are also faster. Because there are more molecules in the gas, the gas pressure is higher and the rate at which gas molecules condense and enter the liquid is faster. As a result the gas and liquid are in equilibrium at this higher temperature.

Substance	**Temperature**		**Pressure**	
	K	ºC	Pa	atm
Water	273.16	0.01	6.10×10^2	0.00600
Carbon dioxide	216.55	−56.60	5.16×10^5	5.11
Sulfur dioxide	197.68	−75.47	1.67×10^3	0.0167
Ammonia	195.40	−77.75	6.06×10^3	0.0600
Nitrogen	63.18	−210.0	1.25×10^4	0.124
Oxygen	54.36	−218.8	1.52×10^2	0.00151
Hydrogen	13.84	−259.3	7.04×10^3	0.0697

Triple Point Temperatures and Pressures

One example of equilibrium between liquid and gas is that of water and steam at 100ºC and 1.00 atm. This temperature is the boiling point at that pressure, so they should exist in equilibrium. Why does an open pot of water at 100ºC boil completely away? The gas surrounding an open pot is not pure water: it is mixed with air. If pure water and steam are in a closed container at $100^\circ C$ and 1.00 atm, they would coexist—but with air over the pot, there are fewer water molecules to condense, and water boils. What about water at $20.0^\circ C$ and 1.00 atm? This temperature and pressure correspond to the liquid region, yet an open glass of water at this temperature will completely evaporate. Again, the gas around it is air and not pure water vapor, so that the reduced evaporation rate is greater than the condensation rate of water from dry air. If the glass is sealed, then the liquid phase remains. We call the gas phase a **vapor** when it exists, as it does for water at $20.0^\circ C$, at a temperature below the boiling temperature.

Exercise:

Check Your Understanding

Problem:

Explain why a cup of water (or soda) with ice cubes stays at $0^\circ C$, even on a hot summer day.

Solution:

The ice and liquid water are in thermal equilibrium, so that the temperature stays at the freezing temperature as long as ice remains in the liquid. (Once all of the ice melts, the water temperature will start to rise.)

Vapor Pressure, Partial Pressure, and Dalton’s Law

Vapor pressure is defined as the pressure at which a gas coexists with its solid or liquid phase. Vapor pressure is created by faster molecules that break away from the liquid or solid and enter the gas phase. The vapor pressure of a substance depends on both the substance and its temperature—an increase in temperature increases the vapor pressure.

Partial pressure is defined as the pressure a gas would create if it occupied the total volume available. In a mixture of gases, *the total pressure is the sum of partial pressures of the component gases*, assuming ideal gas behavior and no chemical reactions between the components. This law is known as **Dalton’s law of partial pressures**, after the English scientist John Dalton (1766–1844), who proposed it. Dalton’s law is based on kinetic theory, where each gas creates its pressure by molecular collisions, independent of other gases present. It is consistent with the fact that pressures add according to Pascal’s Principle. Thus water evaporates and ice sublimates when their vapor pressures exceed the partial pressure of water vapor in the surrounding mixture of gases. If their vapor pressures are less than the partial pressure of water vapor in the surrounding gas, liquid droplets or ice crystals (frost) form.

Exercise:

Check Your Understanding

Problem:

Is energy transfer involved in a phase change? If so, will energy have to be supplied to change phase from solid to liquid and liquid to gas? What about gas to liquid and liquid to solid? Why do they spray the orange trees with water in Florida when the temperatures are near or just below freezing?

Solution:

Yes, energy transfer is involved in a phase change. We know that atoms and molecules in solids and liquids are bound to each other because we know that force is required to separate them. So in a phase change from solid to liquid and liquid to gas, a force must be exerted, perhaps by collision, to separate atoms and molecules. Force exerted through a distance is work, and energy is needed to do work to go from solid to liquid and liquid to gas. This is intuitively consistent with the need for energy to melt ice or boil water. The converse is also true. Going from gas to liquid or liquid to solid involves atoms and molecules pushing together, doing work and releasing energy.

Note:

PhET Explorations: States of Matter—Basics

Heat, cool, and compress atoms and molecules and watch as they change between solid, liquid, and gas phases. https://phet.colorado.edu/sims/html/states-of-matter-basics/latest/states-of-matter-basics_en.html

Section Summary

- Most substances have three distinct phases: gas, liquid, and solid.
- Phase changes among the various phases of matter depend on temperature and pressure.
- The existence of the three phases with respect to pressure and temperature can be described in a phase diagram.
- Two phases coexist (i.e., they are in thermal equilibrium) at a set of pressures and temperatures. These are described as a line on a phase diagram.
- The three phases coexist at a single pressure and temperature. This is known as the triple point and is described by a single point on a phase diagram.
- A gas at a temperature below its boiling point is called a vapor.
- Vapor pressure is the pressure at which a gas coexists with its solid or liquid phase.
- Partial pressure is the pressure a gas would create if it existed alone.
- Dalton’s law states that the total pressure is the sum of the partial pressures of all of the gases present.

Conceptual Questions

Exercise:

Problem:

A pressure cooker contains water and steam in equilibrium at a pressure greater than atmospheric pressure. How does this greater pressure increase cooking speed?

Exercise:

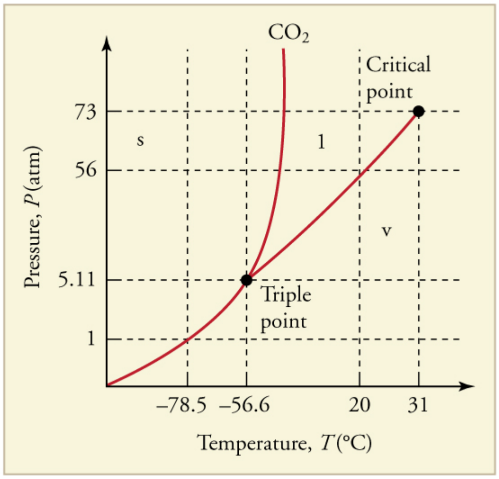
Problem:

Why does condensation form most rapidly on the coldest object in a room—for example, on a glass of ice water?

Exercise:

Problem:

What is the vapor pressure of solid carbon dioxide (dry ice) at $-78.5^{\circ}C$?



The phase diagram for carbon dioxide. The axes are nonlinear, and the graph is not to scale. Dry ice is solid carbon dioxide and has a sublimation temperature of $-78.5^{\circ}C$.

Exercise:

Problem:

Can carbon dioxide be liquefied at room temperature ($20^{\circ}C$)? If so, how? If not, why not? (See [link].)

Exercise:

Problem:

Oxygen cannot be liquefied at room temperature by placing it under a large enough pressure to force its molecules together. Explain why this is.

Exercise:

Problem: What is the distinction between gas and vapor?

Glossary

PV diagram
a graph of pressure vs. volume

critical point
the temperature above which a liquid cannot exist

critical temperature
the temperature above which a liquid cannot exist

critical pressure
the minimum pressure needed for a liquid to exist at the critical temperature

vapor
a gas at a temperature below the boiling temperature

vapor pressure
the pressure at which a gas coexists with its solid or liquid phase

phase diagram
a graph of pressure vs. temperature of a particular substance, showing at which pressures and temperatures the three phases of the substance occur

triple point
the pressure and temperature at which a substance exists in equilibrium as a solid, liquid, and gas

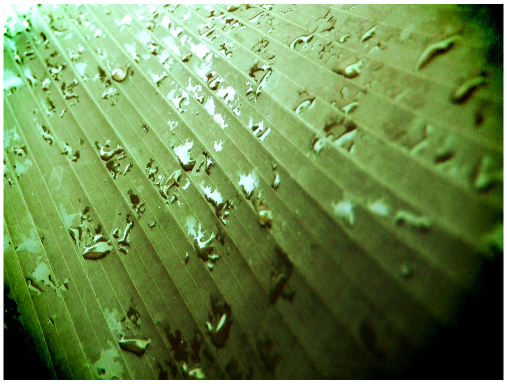
sublimation
the phase change from solid to gas

partial pressure
the pressure a gas would create if it occupied the total volume of space available

Dalton’s law of partial pressures
the physical law that states that the total pressure of a gas is the sum of partial pressures of the component gases

Humidity, Evaporation, and Boiling

- Explain the relationship between vapor pressure of water and the capacity of air to hold water vapor.
- Explain the relationship between relative humidity and partial pressure of water vapor in the air.
- Calculate vapor density using vapor pressure.
- Calculate humidity and dew point.

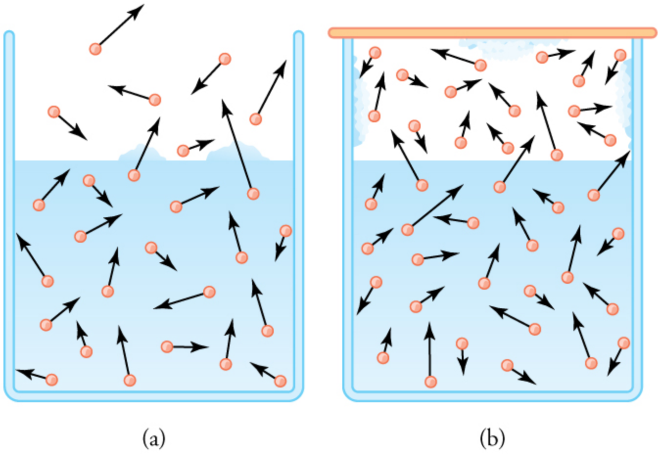


Dew drops like these, on a banana leaf photographed just after sunrise, form when the air temperature drops to or below the dew point. At the dew point, the rate at which water molecules join together is greater than the rate at which they separate, and some of the water condenses to form droplets. (credit: Aaron Escobar, Flickr)

The expression “it’s not the heat, it’s the humidity” makes a valid point. We keep cool in hot weather by evaporating sweat from our skin and water from our breathing passages. Because evaporation is inhibited by high humidity, we feel hotter at a given temperature when the humidity is high. Low humidity, on the other hand, can cause discomfort from excessive drying of mucous membranes and can lead to an increased risk of respiratory infections.

When we say humidity, we really mean **relative humidity**. Relative humidity tells us how much water vapor is in the air compared with the maximum possible. At its maximum, denoted as **saturation**, the relative humidity is 100%, and evaporation is inhibited. The amount of water vapor in the air depends on temperature. For example, relative humidity rises in the evening, as air temperature declines, sometimes reaching the **dew point**. At the dew point temperature, relative humidity is 100%, and fog may result from the condensation of water droplets if they are small enough to stay in suspension. Conversely, if you wish to dry something (perhaps your hair), it is more effective to blow hot air over it rather than cold air, because, among other things, the increase in temperature increases the energy of the molecules, so the rate of evaporation increases.

The amount of water vapor in the air depends on the vapor pressure of water. The liquid and solid phases are continuously giving off vapor because some of the molecules have high enough speeds to enter the gas phase; see [link](a). If a lid is placed over the container, as in [link](b), evaporation continues, increasing the pressure, until sufficient vapor has built up for condensation to balance evaporation. Then equilibrium has been achieved, and the vapor pressure is equal to the partial pressure of water in the container. Vapor pressure increases with temperature because molecular speeds are higher as temperature increases. [link] gives representative values of water vapor pressure over a range of temperatures.



(a) Because of the distribution of speeds and kinetic energies, some water molecules can break away to the vapor phase even at temperatures below the ordinary boiling point. (b) If the container is sealed, evaporation will continue until there is enough vapor density for the condensation rate to equal the evaporation rate. This vapor density and the partial pressure it creates are the saturation values. They increase with temperature and are independent of the presence of other gases, such as air. They depend only on the vapor pressure of water.

Relative humidity is related to the partial pressure of water vapor in the air. At 100% humidity, the partial pressure is equal to the vapor pressure, and no more water can enter the vapor phase. If the partial pressure is less than the vapor pressure, then evaporation will take place, as humidity is less than 100%. If the partial pressure is greater than the vapor pressure, condensation takes place. In everyday language, people sometimes refer to the capacity of air to “hold” water vapor, but this is not actually what happens. The water vapor is not held by the air. The amount of water in air is determined by the vapor pressure of water and has nothing to do with the properties of air.

Temperature ($^{\circ}$C)	**Vapor pressure (Pa)**	**Saturation vapor density (g/m^3)**
−50	4.0	0.039
−20	1.04×10^2	0.89
−10	2.60×10^2	2.36
0	6.10×10^2	4.84
5	8.68×10^2	6.80
10	1.19×10^3	9.40

Temperature (ºC)	**Vapor pressure (Pa)**	**Saturation vapor density (g/m^3)**
15	1.69×10^3	12.8
20	2.33×10^3	17.2
25	3.17×10^3	23.0
30	4.24×10^3	30.4
37	6.31×10^3	44.0
40	7.34×10^3	51.1
50	1.23×10^4	82.4
60	1.99×10^4	130
70	3.12×10^4	197

Temperature $(^\circ\text{C})$	**Vapor pressure (Pa)**	**Saturation vapor density (g/m^3)**
80	4.73×10^4	294
90	7.01×10^4	418
95	8.59×10^4	505
100	$\mathbf{1.01 \times 10^5}$	**598**
120	1.99×10^5	1095
150	4.76×10^5	2430
200	1.55×10^6	7090
220	2.32×10^6	10,200

Saturation Vapor Density of Water

Example:

Calculating Density Using Vapor Pressure

[link] gives the vapor pressure of water at 20.0°C as $2.33 \times 10^3\ \text{Pa}$. Use the ideal gas law to calculate the density of water vapor in g/m^3 that would create a partial pressure equal to this vapor pressure. Compare the result with the saturation vapor density given in the table.

Strategy

To solve this problem, we need to break it down into a two steps. The partial pressure follows the ideal gas law,

Equation:

$$\text{PV} = \text{nRT},$$

where n is the number of moles. If we solve this equation for n/V to calculate the number of moles per cubic meter, we can then convert this quantity to grams per cubic meter as requested. To do this, we need to use the molecular mass of water, which is given in the periodic table.

Solution

1. Identify the knowns and convert them to the proper units:

   a. temperature $T = 20^{\circ}\text{C}{=}293\ \text{K}$
   b. vapor pressure P of water at 20°C is $2.33 \times 10^3\ \text{Pa}$
   c. molecular mass of water is $18.0\ \text{g/mol}$

2. Solve the ideal gas law for n/V.

Equation:

$$\frac{n}{V} = \frac{P}{\text{RT}}$$

3. Substitute known values into the equation and solve for n/V.

Equation:

$$\frac{n}{V} = \frac{P}{\text{RT}} = \frac{2.33 \times 10^3\ \text{Pa}}{(8.31\ \text{J/mol} \cdot \text{K})(293\ \text{K})} = 0.957\ \text{mol/m}^3$$

4. Convert the density in moles per cubic meter to grams per cubic meter.

Equation:

$$\rho = \left(0.957\frac{\text{mol}}{\text{m}^3}\right)\left(\frac{18.0\ \text{g}}{\text{mol}}\right) = 17.2\ \text{g/m}^3$$

Discussion

The density is obtained by assuming a pressure equal to the vapor pressure of water at 20.0ºC. The density found is identical to the value in [link], which means that a vapor density of $17.2\ \text{g/m}^3$ at 20.0ºC creates a partial pressure of $2.33 \times 10^3\ \text{Pa}$, equal to the vapor pressure of water at that temperature. If the partial pressure is equal to the vapor pressure, then the liquid and vapor phases are in equilibrium, and the relative humidity is 100%. Thus, there can be no more than 17.2 g of water vapor per m^3 at 20.0ºC, so that this value is the saturation vapor density at that temperature. This example illustrates how water vapor behaves like an ideal gas: the pressure and density are consistent with the ideal gas law (assuming the density in the table is correct). The saturation vapor densities listed in [link] are the maximum amounts of water vapor that air can hold at various temperatures.

Note:

Percent Relative Humidity

We define **percent relative humidity** as the ratio of vapor density to saturation vapor density, or

Equation:

$$\text{percent relative humidity} = \frac{\text{vapor density}}{\text{saturation vapor density}} \times 100$$

We can use this and the data in [link] to do a variety of interesting calculations, keeping in mind that relative humidity is based on the comparison of the partial pressure of water vapor in air and ice.

Example:

Calculating Humidity and Dew Point

(a) Calculate the percent relative humidity on a day when the temperature is 25.0°C and the air contains 9.40 g of water vapor per m^3. (b) At what temperature will this air reach 100% relative humidity (the saturation density)? This temperature is the dew point. (c) What is the humidity when the air temperature is 25.0°C and the dew point is -10.0°C?

Strategy and Solution

(a) Percent relative humidity is defined as the ratio of vapor density to saturation vapor density.

Equation:

$$\text{percent relative humidity} = \frac{\text{vapor density}}{\text{saturation vapor density}} \times 100$$

The first is given to be $9.40\ \text{g/m}^3$, and the second is found in [link] to be $23.0\ \text{g/m}^3$. Thus,

Equation:

$$\text{percent relative humidity} = \frac{9.40\ \text{g/m}^3}{23.0\ \text{g/m}^3} \times 100 = 40.9.\%$$

(b) The air contains $9.40\ \text{g/m}^3$ of water vapor. The relative humidity will be 100% at a temperature where $9.40\ \text{g/m}^3$ is the saturation density. Inspection of [link] reveals this to be the case at 10.0°C, where the relative humidity will be 100%. That temperature is called the dew point for air with this concentration of water vapor.

(c) Here, the dew point temperature is given to be -10.0°C. Using [link], we see that the vapor density is $2.36\ \text{g/m}^3$, because this value is the saturation vapor density at -10.0°C. The saturation vapor density at 25.0°C is seen to be $23.0\ \text{g/m}^3$. Thus, the relative humidity at 25.0°C is

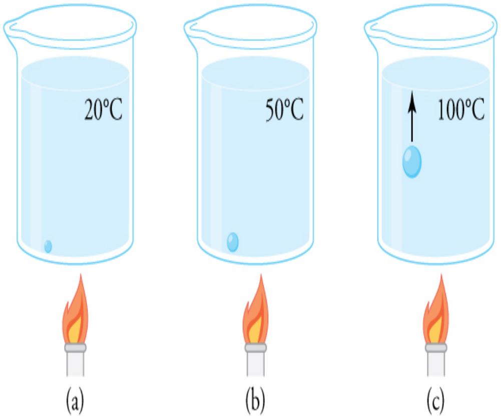
Equation:

$$\text{percent relative humidity} = \frac{2.36\ \text{g/m}^3}{23.0\ \text{g/m}^3} \times 100 = 10.3\%.$$

Discussion

The importance of dew point is that air temperature cannot drop below $10.0^{\circ}\mathrm{C}$ in part (b), or $-10.0^{\circ}\mathrm{C}$ in part (c), without water vapor condensing out of the air. If condensation occurs, considerable transfer of heat occurs (discussed in Heat and Heat Transfer Methods), which prevents the temperature from further dropping. When dew points are below $0^{\circ}\mathrm{C}$, freezing temperatures are a greater possibility, which explains why farmers keep track of the dew point. Low humidity in deserts means low dew-point temperatures. Thus condensation is unlikely. If the temperature drops, vapor does not condense in liquid drops. Because no heat is released into the air, the air temperature drops more rapidly compared to air with higher humidity. Likewise, at high temperatures, liquid droplets do not evaporate, so that no heat is removed from the gas to the liquid phase. This explains the large range of temperature in arid regions.

Why does water boil at $100^{\circ}\mathrm{C}$? You will note from [link] that the vapor pressure of water at $100^{\circ}\mathrm{C}$ is $1.01 \times 10^{5}\ \mathrm{Pa}$, or 1.00 atm. Thus, it can evaporate without limit at this temperature and pressure. But why does it form bubbles when it boils? This is because water ordinarily contains significant amounts of dissolved air and other impurities, which are observed as small bubbles of air in a glass of water. If a bubble starts out at the bottom of the container at $20^{\circ}\mathrm{C}$, it contains water vapor (about 2.30%). The pressure inside the bubble is fixed at 1.00 atm (we ignore the slight pressure exerted by the water around it). As the temperature rises, the amount of air in the bubble stays the same, but the water vapor increases; the bubble expands to keep the pressure at 1.00 atm. At $100^{\circ}\mathrm{C}$, water vapor enters the bubble continuously since the partial pressure of water is equal to 1.00 atm in equilibrium. It cannot reach this pressure, however, since the bubble also contains air and total pressure is 1.00 atm. The bubble grows in size and thereby increases the buoyant force. The bubble breaks away and rises rapidly to the surface—we call this boiling! (See [link].)



(a) An air bubble in water starts out saturated with water vapor at 20°C. (b) As the temperature rises, water vapor enters the bubble because its vapor pressure increases. The bubble expands to keep its pressure at 1.00 atm. (c) At 100°C, water vapor enters the bubble continuously because water’s vapor pressure exceeds its partial pressure in the bubble, which must be less than 1.00 atm. The bubble grows and rises to the surface.

Exercise:
Check Your Understanding

Problem:

Freeze drying is a process in which substances, such as foods, are dried by placing them in a vacuum chamber and lowering the atmospheric pressure around them. How does the lowered atmospheric pressure speed the drying process, and why does it cause the temperature of the food to drop?

Solution:

Decreased the atmospheric pressure results in decreased partial pressure of water, hence a lower humidity. So evaporation of water from food, for example, will be enhanced. The molecules of water most likely to break away from the food will be those with the greatest velocities. Those remaining thus have a lower average velocity and a lower temperature. This can (and does) result in the freezing and drying of the food; hence the process is aptly named freeze drying.

Note:
PhET Explorations: States of Matter
Watch different types of molecules form a solid, liquid, or gas. Add or remove heat and watch the phase change. Change the temperature or volume of a container and see a pressure-temperature diagram respond in real time. Relate the interaction potential to the forces between molecules.
https://phet.colorado.edu/sims/html/states-of-matter/latest/states-of-matter_en.html

Section Summary

- Relative humidity is the fraction of water vapor in a gas compared to the saturation value.
- The saturation vapor density can be determined from the vapor pressure for a given temperature.

- Percent relative humidity is defined to be
  Equation:

$$\text{percent relative humidity} = \frac{\text{vapor density}}{\text{saturation vapor density}} \times 100.$$

- The dew point is the temperature at which air reaches 100% relative humidity.

Conceptual Questions

Exercise:

Problem:

Because humidity depends only on water's vapor pressure and temperature, are the saturation vapor densities listed in [link] valid in an atmosphere of helium at a pressure of $1.01 \times 10^5 \text{ N/m}^2$, rather than air? Are those values affected by altitude on Earth?

Exercise:

Problem:

Why does a beaker of 40.0°C water placed in a vacuum chamber start to boil as the chamber is evacuated (air is pumped out of the chamber)? At what pressure does the boiling begin? Would food cook any faster in such a beaker?

Exercise:

Problem:

Why does rubbing alcohol evaporate much more rapidly than water at STP (standard temperature and pressure)?

Problems & Exercises

Exercise:

Problem:

Dry air is 78.1% nitrogen. What is the partial pressure of nitrogen when the atmospheric pressure is $1.01 \times 10^5\ \mathrm{N/m^2}$?

Solution:

$7.89 \times 10^4\ \mathrm{Pa}$

Exercise:

Problem:

(a) What is the vapor pressure of water at $20.0^{\mathrm{o}}\mathrm{C}$? (b) What percentage of atmospheric pressure does this correspond to? (c) What percent of $20.0^{\mathrm{o}}\mathrm{C}$ air is water vapor if it has 100% relative humidity? (The density of dry air at $20.0^{\mathrm{o}}\mathrm{C}$ is $1.20\ \mathrm{kg/m^3}$.)

Exercise:

Problem:

Pressure cookers increase cooking speed by raising the boiling temperature of water above its value at atmospheric pressure. (a) What pressure is necessary to raise the boiling point to $120.0^{\mathrm{o}}\mathrm{C}$? (b) What gauge pressure does this correspond to?

Solution:

(a) $1.99 \times 10^5\ \mathrm{Pa}$

(b) 0.97 atm

Exercise:

Problem:

(a) At what temperature does water boil at an altitude of 1500 m (about 5000 ft) on a day when atmospheric pressure is $8.59 \times 10^{4}\ \text{N/m}^{2}$? (b) What about at an altitude of 3000 m (about 10,000 ft) when atmospheric pressure is $7.00 \times 10^{4}\ \text{N/m}^{2}$?

Exercise:

Problem:

What is the atmospheric pressure on top of Mt. Everest on a day when water boils there at a temperature of 70.0°C?

Solution:

$3.12 \times 10^{4}\ \text{Pa}$

Exercise:

Problem:

At a spot in the high Andes, water boils at 80.0°C, greatly reducing the cooking speed of potatoes, for example. What is atmospheric pressure at this location?

Exercise:

Problem:

What is the relative humidity on a 25.0°C day when the air contains $18.0\ \text{g/m}^{3}$ of water vapor?

Solution:

78.3%

Exercise:

Problem:

What is the density of water vapor in $\mathrm{g/m^3}$ on a hot dry day in the desert when the temperature is 40.0ºC and the relative humidity is 6.00%?

Exercise:

Problem:

A deep-sea diver should breathe a gas mixture that has the same oxygen partial pressure as at sea level, where dry air contains 20.9% oxygen and has a total pressure of $1.01 \times 10^5\ \mathrm{N/m^2}$. (a) What is the partial pressure of oxygen at sea level? (b) If the diver breathes a gas mixture at a pressure of $2.00 \times 10^6\ \mathrm{N/m^2}$, what percent oxygen should it be to have the same oxygen partial pressure as at sea level?

Solution:

(a) $2.12 \times 10^4\ \mathrm{Pa}$

(b) $1.06\ \%$

Exercise:

Problem:

The vapor pressure of water at 40.0ºC is $7.34 \times 10^3\ \mathrm{N/m^2}$. Using the ideal gas law, calculate the density of water vapor in $\mathrm{g/m^3}$ that creates a partial pressure equal to this vapor pressure. The result should be the same as the saturation vapor density at that temperature $(51.1\ \mathrm{g/m^3})$.

Exercise:

Problem:

Air in human lungs has a temperature of $37.0^{\mathrm{o}}\mathrm{C}$ and a saturation vapor density of $44.0\ \mathrm{g/m}^3$. (a) If 2.00 L of air is exhaled and very dry air inhaled, what is the maximum loss of water vapor by the person? (b) Calculate the partial pressure of water vapor having this density, and compare it with the vapor pressure of $6.31 \times 10^3\ \mathrm{N/m}^2$.

Solution:

(a) $8.80 \times 10^{-2}\ \mathrm{g}$

(b) $6.30 \times 10^3\ \mathrm{Pa}$; the two values are nearly identical.

Exercise:

Problem:

If the relative humidity is 90.0% on a muggy summer morning when the temperature is $20.0^{\mathrm{o}}\mathrm{C}$, what will it be later in the day when the temperature is $30.0^{\mathrm{o}}\mathrm{C}$, assuming the water vapor density remains constant?

Exercise:

Problem:

Late on an autumn day, the relative humidity is 45.0% and the temperature is $20.0^{\mathrm{o}}\mathrm{C}$. What will the relative humidity be that evening when the temperature has dropped to $10.0^{\mathrm{o}}\mathrm{C}$, assuming constant water vapor density?

Solution:

82.3%

Exercise:

Problem:

Atmospheric pressure atop Mt. Everest is $3.30\times10^4\ \text{N/m}^2$. (a) What is the partial pressure of oxygen there if it is 20.9% of the air? (b) What percent oxygen should a mountain climber breathe so that its partial pressure is the same as at sea level, where atmospheric pressure is $1.01\times10^5\ \text{N/m}^2$? (c) One of the most severe problems for those climbing very high mountains is the extreme drying of breathing passages. Why does this drying occur?

Exercise:

Problem:

What is the dew point (the temperature at which 100% relative humidity would occur) on a day when relative humidity is 39.0% at a temperature of 20.0^{o}C?

Solution:

4.77^{o}C

Exercise:

Problem:

On a certain day, the temperature is 25.0^{o}C and the relative humidity is 90.0%. How many grams of water must condense out of each cubic meter of air if the temperature falls to 15.0^{o}C? Such a drop in temperature can, thus, produce heavy dew or fog.

Exercise:

Problem: Integrated Concepts

The boiling point of water increases with depth because pressure increases with depth. At what depth will fresh water have a boiling point of 150^{o}C, if the surface of the water is at sea level?

Solution:

$38.3\ \text{m}$

Exercise:

Problem: Integrated Concepts

(a) At what depth in fresh water is the critical pressure of water reached, given that the surface is at sea level? (b) At what temperature will this water boil? (c) Is a significantly higher temperature needed to boil water at a greater depth?

Exercise:

Problem: Integrated Concepts

To get an idea of the small effect that temperature has on Archimedes' principle, calculate the fraction of a copper block's weight that is supported by the buoyant force in 0^{o}C water and compare this fraction with the fraction supported in 95.0^{o}C water.

Solution:

$\frac{(F_{\text{B}}/w_{\text{Cu}})}{(F_{\text{B}}/w_{\text{Cu}})'} = 1.02$. The buoyant force supports nearly the exact same amount of force on the copper block in both circumstances.

Exercise:

Problem: Integrated Concepts

If you want to cook in water at 150^{o}C, you need a pressure cooker that can withstand the necessary pressure. (a) What pressure is required for the boiling point of water to be this high? (b) If the lid of the pressure cooker is a disk 25.0 cm in diameter, what force must it be able to withstand at this pressure?

Exercise:

Problem: **Unreasonable Results**

(a) How many moles per cubic meter of an ideal gas are there at a pressure of $1.00 \times 10^{14}\ \mathrm{N/m^2}$ and at $0^{\mathrm{o}}\mathrm{C}$? (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Solution:

(a) $4.41 \times 10^{10}\ \mathrm{mol/m^3}$

(b) It’s unreasonably large.

(c) At high pressures such as these, the ideal gas law can no longer be applied. As a result, unreasonable answers come up when it is used.

Exercise:

Problem: **Unreasonable Results**

(a) An automobile mechanic claims that an aluminum rod fits loosely into its hole on an aluminum engine block because the engine is hot and the rod is cold. If the hole is 10.0% bigger in diameter than the $22.0^{\mathrm{o}}\mathrm{C}$ rod, at what temperature will the rod be the same size as the hole? (b) What is unreasonable about this temperature? (c) Which premise is responsible?

Exercise:

Problem: **Unreasonable Results**

The temperature inside a supernova explosion is said to be $2.00 \times 10^{13}\ \mathrm{K}$. (a) What would the average velocity v_{rms} of hydrogen atoms be? (b) What is unreasonable about this velocity? (c) Which premise or assumption is responsible?

Solution:

(a) $7.03\times 10^{8}\text{ m/s}$

(b) The velocity is too high—it's greater than the speed of light.

(c) The assumption that hydrogen inside a supernova behaves as an idea gas is responsible, because of the great temperature and density in the core of a star. Furthermore, when a velocity greater than the speed of light is obtained, classical physics must be replaced by relativity, a subject not yet covered.

Exercise:

Problem: Unreasonable Results

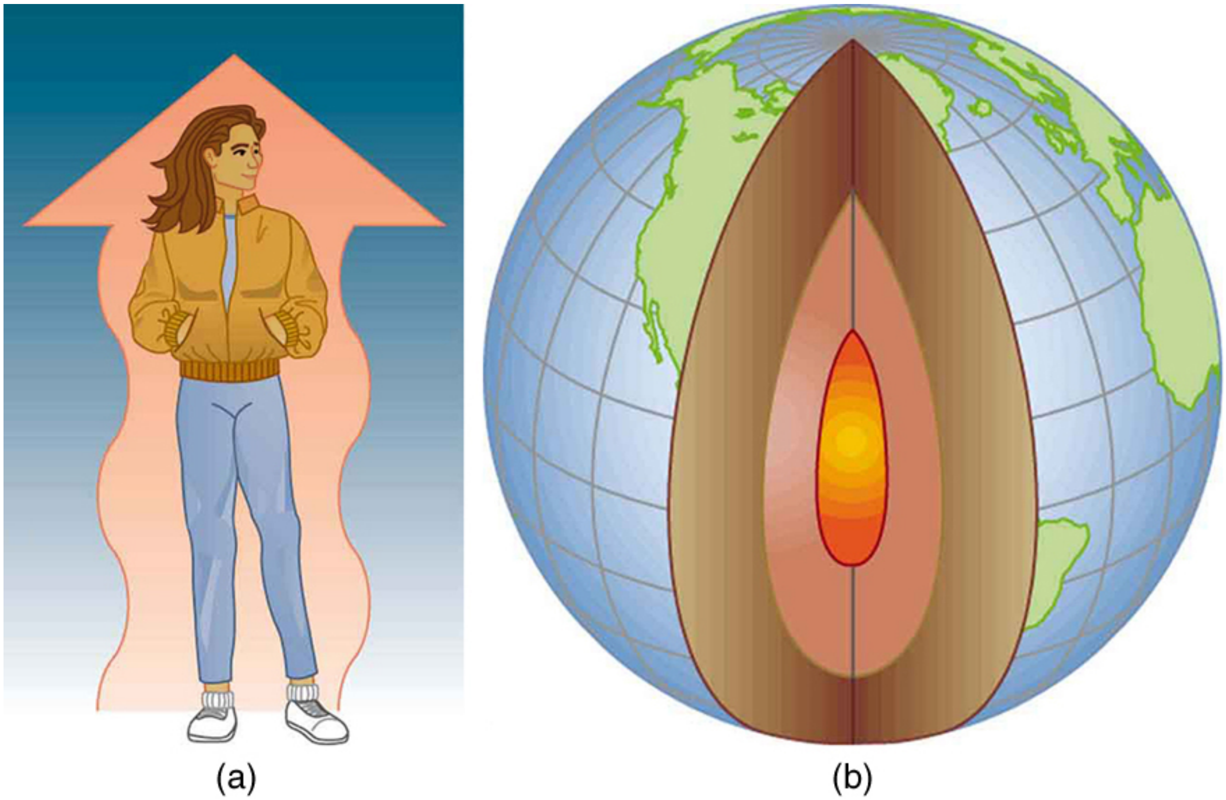
Suppose the relative humidity is 80% on a day when the temperature is 30.0°C. (a) What will the relative humidity be if the air cools to 25.0°C and the vapor density remains constant? (b) What is unreasonable about this result? (c) Which premise is responsible?

Glossary

dew point	the temperature at which relative humidity is 100%; the temperature at which water starts to condense out of the air
saturation	the condition of 100% relative humidity
percent relative humidity	the ratio of vapor density to saturation vapor density
relative humidity	the amount of water in the air relative to the maximum amount the air can hold

Introduction to Heat and Heat Transfer Methods class="introduction"

(a) The chilling effect of a clear breezy night is produced by the wind and by radiative heat transfer to cold outer space. (b) There was once great controversy about the Earth’s age, but it is now generally accepted to be about 4.5 billion years old. Much of the debate is centered on the Earth’s molten interior. According to our understanding of heat transfer, if the Earth is really that old, its center should have cooled off long ago. The discovery of radioactivity in rocks revealed the source of energy that keeps the Earth’s interior molten, despite heat transfer to the surface, and from there to cold outer space.



Energy can exist in many forms and heat is one of the most intriguing. Heat is often hidden, as it only exists when in transit, and is transferred by a number of distinctly different methods. Heat transfer touches every aspect of our lives and helps us understand how the universe functions. It explains the chill we feel on a clear breezy night, or why Earth's core has yet to cool. This chapter defines and explores heat transfer, its effects, and the methods by which heat is transferred. These topics are fundamental, as well as practical, and will often be referred to in the chapters ahead.

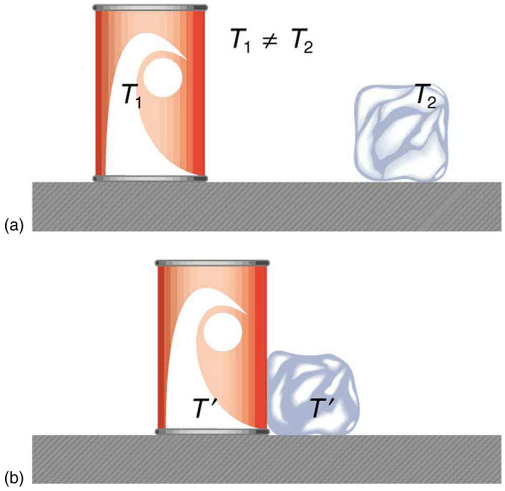
Heat

- Define heat as transfer of energy.

In Work, Energy, and Energy Resources, we defined work as force times distance and learned that work done on an object changes its kinetic energy. We also saw in Temperature, Kinetic Theory, and the Gas Laws that temperature is proportional to the (average) kinetic energy of atoms and molecules. We say that a thermal system has a certain internal energy: its internal energy is higher if the temperature is higher. If two objects at different temperatures are brought in contact with each other, energy is transferred from the hotter to the colder object until equilibrium is reached and the bodies reach thermal equilibrium (i.e., they are at the same temperature). No work is done by either object, because no force acts through a distance. The transfer of energy is caused by the temperature difference, and ceases once the temperatures are equal. These observations lead to the following definition of **heat**: Heat is the spontaneous transfer of energy due to a temperature difference.

As noted in Temperature, Kinetic Theory, and the Gas Laws, heat is often confused with temperature. For example, we may say the heat was unbearable, when we actually mean that the temperature was high. Heat is a form of energy, whereas temperature is not. The misconception arises because we are sensitive to the flow of heat, rather than the temperature.

Owing to the fact that heat is a form of energy, it has the SI unit of *joule* (J). The *calorie* (cal) is a common unit of energy, defined as the energy needed to change the temperature of 1.00 g of water by 1.00ºC —specifically, between 14.5ºC and 15.5ºC, since there is a slight temperature dependence. Perhaps the most common unit of heat is the **kilocalorie** (kcal), which is the energy needed to change the temperature of 1.00 kg of water by 1.00ºC. Since mass is most often specified in kilograms, kilocalorie is commonly used. Food calories (given the notation Cal, and sometimes called “big calorie”) are actually kilocalories (1 kilocalorie = 1000 calories), a fact not easily determined from package labeling.



In figure (a) the soft drink and the ice have different temperatures, T_1 and T_2, and are not in thermal equilibrium. In figure (b), when the soft drink and ice are allowed to interact, energy is transferred until they reach the same temperature T', achieving equilibrium. Heat transfer occurs due to the difference in temperatures. In fact, since the soft drink and ice are both in contact with the surrounding air and bench, the equilibrium temperature will be the same for both.

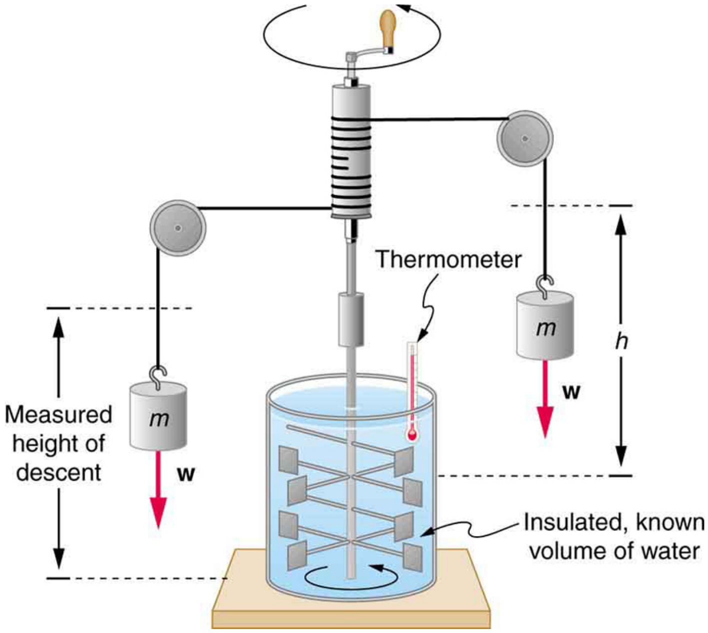
Mechanical Equivalent of Heat

It is also possible to change the temperature of a substance by doing work. Work can transfer energy into or out of a system. This realization helped establish the fact that heat is a form of energy. James Prescott Joule (1818–1889) performed many experiments to establish the **mechanical equivalent of heat**—*the work needed to produce the same effects as heat transfer*. In terms of the units used for these two terms, the best modern value for this equivalence is

Equation:

$$1.000\ \text{kcal} = 4186\ \text{J}.$$

We consider this equation as the conversion between two different units of energy.


Thermometer
m
h
w
Measured height of descent
m
w
Insulated, known volume of water

Schematic depiction of Joule's experiment that established the equivalence of heat and work.

The figure above shows one of Joule’s most famous experimental setups for demonstrating the mechanical equivalent of heat. It demonstrated that work and heat can produce the same effects, and helped establish the principle of conservation of energy. Gravitational potential energy (PE) (work done by the gravitational force) is converted into kinetic energy (KE), and then randomized by viscosity and turbulence into increased average kinetic energy of atoms and molecules in the system, producing a temperature increase. His contributions to the field of thermodynamics were so significant that the SI unit of energy was named after him.

Heat added or removed from a system changes its internal energy and thus its temperature. Such a temperature increase is observed while cooking. However, adding heat does not necessarily increase the temperature. An example is melting of ice; that is, when a substance changes from one phase to another. Work done on the system or by the system can also change the internal energy of the system. Joule demonstrated that the temperature of a system can be increased by stirring. If an ice cube is rubbed against a rough surface, work is done by the frictional force. A system has a well-defined internal energy, but we cannot say that it has a certain “heat content” or “work content”. We use the phrase “heat transfer” to emphasize its nature.

Exercise:

Check Your Understanding

Problem:

Two samples (A and B) of the same substance are kept in a lab. Someone adds 10 kilojoules (kJ) of heat to one sample, while 10 kJ of work is done on the other sample. How can you tell to which sample the heat was added?

Solution:

Heat and work both change the internal energy of the substance. However, the properties of the sample only depend on the internal energy so that it is impossible to tell whether heat was added to sample A or B.

Summary

- Heat and work are the two distinct methods of energy transfer.
- Heat is energy transferred solely due to a temperature difference.
- Any energy unit can be used for heat transfer, and the most common are kilocalorie (kcal) and joule (J).
- Kilocalorie is defined to be the energy needed to change the temperature of 1.00 kg of water between 14.5°C and 15.5°C.
- The mechanical equivalent of this heat transfer is $1.00\text{ kcal } = \text{ 4186 J}$.

Conceptual Questions

Exercise:

Problem: How is heat transfer related to temperature?

Exercise:

Problem:

Describe a situation in which heat transfer occurs. What are the resulting forms of energy?

Exercise:

Problem:

When heat transfers into a system, is the energy stored as heat? Explain briefly.

Glossary

heat
the spontaneous transfer of energy due to a temperature difference

kilocalorie
$1\text{ kilocalorie} = 1000\text{ calories}$

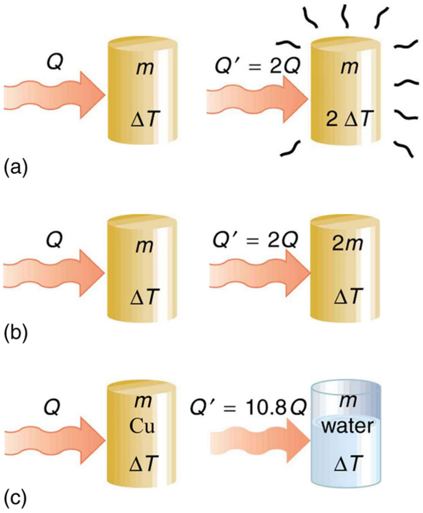
mechanical equivalent of heat

the work needed to produce the same effects as heat transfer

Temperature Change and Heat Capacity

- Observe heat transfer and change in temperature and mass.
- Calculate final temperature after heat transfer between two objects.

One of the major effects of heat transfer is temperature change: heating increases the temperature while cooling decreases it. We assume that there is no phase change and that no work is done on or by the system. Experiments show that the transferred heat depends on three factors—the change in temperature, the mass of the system, and the substance and phase of the substance.



The heat Q transferred to cause a temperature change depends on the magnitude of the temperature change, the mass of the system, and the substance and phase involved. (a) The amount of heat transferred is directly proportional to the temperature change. To double the temperature change of a mass m, you need to add twice the heat. (b) The amount of heat transferred is also directly proportional to the mass. To cause an equivalent temperature change in a doubled mass, you need to add twice the heat. (c) The amount of heat transferred depends on the substance and its phase. If it takes an amount Q of heat to cause a temperature change ΔT in a given mass of copper, it will take 10.8 times that amount of heat to cause the equivalent temperature change in the same mass of water assuming no phase change in either substance.

The dependence on temperature change and mass are easily understood. Owing to the fact that the (average) kinetic energy of an atom or molecule is proportional to the absolute temperature, the internal energy of a system is proportional to the absolute temperature and the number of atoms or molecules. Owing to the fact that the transferred heat is equal to the change in the internal energy, the heat is proportional to the mass of the substance and the temperature change. The transferred heat also depends on the substance so that, for example, the heat necessary to raise the temperature is less for alcohol than for water. For the same substance, the transferred heat also depends on the phase (gas, liquid, or solid).

Note:
Heat Transfer and Temperature Change
The quantitative relationship between heat transfer and temperature change contains all three factors:
Equation:

$$Q = \mathrm{mc}\Delta T,$$

where Q is the symbol for heat transfer, m is the mass of the substance, and ΔT is the change in temperature. The symbol c stands for **specific heat** and depends on the material and phase. The specific heat is the amount of heat necessary to change the temperature of 1.00 kg of mass by 1.00°C. The specific heat c is a property of the substance; its SI unit is $\text{J}/(\text{kg}\cdot\text{K})$ or $\text{J}/(\text{kg}\cdot{}^{\circ}\text{C})$. Recall that the temperature change (ΔT) is the same in units of kelvin and degrees Celsius. If heat transfer is measured in kilocalories, then *the unit of specific heat* is $\text{kcal}/(\text{kg}\cdot{}^{\circ}\text{C})$.

Values of specific heat must generally be looked up in tables, because there is no simple way to calculate them. In general, the specific heat also depends on the temperature. [link] lists representative values of specific heat for various substances. Except for gases, the temperature and volume dependence of the specific heat of most substances is weak. We see from this table that the specific heat of water is five times that of glass and ten times that of iron, which means that it takes five times as much heat to raise the temperature of water the same amount as for glass and ten times as much heat to raise the temperature of water as for iron. In fact, water has one of the largest specific heats of any material, which is important for sustaining life on Earth.

Example:
Calculating the Required Heat: Heating Water in an Aluminum Pan
A 0.500 kg aluminum pan on a stove is used to heat 0.250 liters of water from 20.0°C to 80.0°C. (a) How much heat is required? What percentage of the heat is used to raise the temperature of (b) the pan and (c) the water?

Strategy
The pan and the water are always at the same temperature. When you put the pan on the stove, the temperature of the water and the pan is increased by the same amount. We use the equation for the heat transfer for the given temperature change and mass of water and aluminum. The specific heat values for water and aluminum are given in [link].

Solution
Because water is in thermal contact with the aluminum, the pan and the water are at the same temperature.

1. Calculate the temperature difference:
   Equation:
   $$\Delta T = T_{\text{f}} - T_{\text{i}} = 60.0^{\circ}\text{C}.$$
2. Calculate the mass of water. Because the density of water is $1000\ \text{kg}/\text{m}^3$, one liter of water has a mass of 1 kg, and the mass of 0.250 liters of water is $m_{\text{w}} = 0.250\ \text{kg}$.
3. Calculate the heat transferred to the water. Use the specific heat of water in [link]:
   Equation:

$$Q_{\text{w}} = m_{\text{w}} c_{\text{w}} \Delta T = (0.250\ \text{kg})(4186\ \text{J/kg}^{\text{o}}\text{C})(60.0^{\text{o}}\text{C}) = 62.8\ \text{kJ}.$$

4. Calculate the heat transferred to the aluminum. Use the specific heat for aluminum in [link]:
   Equation:

   $$Q_{\text{Al}} = m_{\text{Al}} c_{\text{Al}} \Delta T = (0.500\ \text{kg})(900\ \text{J/kg}^{\text{o}}\text{C})(60.0^{\text{o}}\text{C}) = 27.0 \times 10^{4}\text{J} = 27.0\ \text{kJ}.$$

5. Compare the percentage of heat going into the pan versus that going into the water. First, find the total transferred heat:
   Equation:

   $$Q_{\text{Total}} = Q_{\text{W}} + Q_{\text{Al}} = 62.8\ \text{kJ} + 27.0\ \ \text{kJ} = 89.8\ \text{kJ}.$$

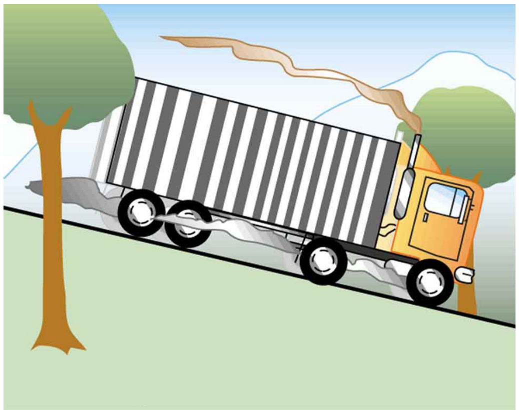
Thus, the amount of heat going into heating the pan is
Equation:

$$\frac{27.0\ \text{kJ}}{89.8\ \text{kJ}} \times 100\% = 30.1\%,$$

and the amount going into heating the water is
Equation:

$$\frac{62.8\ \text{kJ}}{89.8\ \text{kJ}} \times 100\% = 69.9\%.$$

Discussion
In this example, the heat transferred to the container is a significant fraction of the total transferred heat. Although the mass of the pan is twice that of the water, the specific heat of water is over four times greater than that of aluminum. Therefore, it takes a bit more than twice the heat to achieve the given temperature change for the water as compared to the aluminum pan.



The smoking brakes on this truck are a visible evidence of the mechanical equivalent of heat.

Example:

Calculating the Temperature Increase from the Work Done on a Substance: Truck Brakes Overheat on Downhill Runs

Truck brakes used to control speed on a downhill run do work, converting gravitational potential energy into increased internal energy (higher temperature) of the brake material. This conversion prevents the gravitational potential energy from being converted into kinetic energy of the truck. The problem is that the mass of the truck is large compared with that of the brake material absorbing the energy, and the temperature increase may occur too fast for sufficient heat to transfer from the brakes to the environment.

Calculate the temperature increase of 100 kg of brake material with an average specific heat of $800\ \text{J/kg} \cdot {}^{\text{o}}\text{C}$ if the material retains 10% of the energy from a 10,000-kg truck descending 75.0 m (in vertical displacement) at a constant speed.

Strategy

If the brakes are not applied, gravitational potential energy is converted into kinetic energy. When brakes are applied, gravitational potential energy is converted into internal energy of the brake material. We first calculate the gravitational potential energy (Mgh) that the entire truck loses in its descent and then find the temperature increase produced in the brake material alone.

Solution

1. Calculate the change in gravitational potential energy as the truck goes downhill
   Equation:

$$\mathrm{Mgh} = (10{,}000\ \mathrm{kg})\left(9.80\ \mathrm{m/s}^2\right)(75.0\ \mathrm{m}) = 7.35 \times 10^6\ \mathrm{J}.$$

2. Calculate the temperature from the heat transferred using Q=Mgh and
   Equation:

$$\Delta T = \frac{Q}{\mathrm{mc}},$$

where m is the mass of the brake material. Insert the values $m = 100\ \mathrm{kg}$ and $c = 800\ \mathrm{J/kg} \cdot {}^{\mathrm{o}}\mathrm{C}$ to find
Equation:

$$\Delta T = \frac{\left(7.35 \times 10^5\ \mathrm{J}\right)}{(100\ \mathrm{kg})(800\ \mathrm{J/kg^oC})} = 9.2^{\mathrm{o}}\mathrm{C}.$$

Discussion
This same idea underlies the recent hybrid technology of cars, where mechanical energy (gravitational potential energy) is converted by the brakes into electrical energy (battery).

Substances	**Specific heat (*c*)**	
Solids	J/kg·°C	kcal/kg·°C[footnote] These values are identical in units of $\mathrm{cal/g}$ ·°C.
Aluminum	900	0.215
Asbestos	800	0.19
Concrete, granite (average)	840	0.20
Copper	387	0.0924
Glass	840	0.20

Substances	**Specific heat (c)**	
Gold	129	0.0308
Human body (average at 37 °C)	3500	0.83
Ice (average, -50°C to 0°C)	2090	0.50
Iron, steel	452	0.108
Lead	128	0.0305
Silver	235	0.0562
Wood	1700	0.4
Liquids		
Benzene	1740	0.415
Ethanol	2450	0.586
Glycerin	2410	0.576
Mercury	139	0.0333
Water (15.0 °C)	4186	1.000
Gases [footnote] c_v at constant volume and at 20.0ºC, except as noted, and at 1.00 atm average pressure. Values in parentheses are c_p at a constant pressure of 1.00 atm.		
Air (dry)	721 (1015)	0.172 (0.242)
Ammonia	1670 (2190)	0.399 (0.523)
Carbon dioxide	638 (833)	0.152 (0.199)

Substances	**Specific heat (*c*)**	
Nitrogen	739 (1040)	0.177 (0.248)
Oxygen	651 (913)	0.156 (0.218)
Steam (100°C)	1520 (2020)	0.363 (0.482)

Specific Heats[footnote] of Various Substances
The values for solids and liquids are at constant volume and at 25°C, except as noted.

Note that [link] is an illustration of the mechanical equivalent of heat. Alternatively, the temperature increase could be produced by a blow torch instead of mechanically.

Example:
Calculating the Final Temperature When Heat Is Transferred Between Two Bodies: Pouring Cold Water in a Hot Pan
Suppose you pour 0.250 kg of 20.0°C water (about a cup) into a 0.500-kg aluminum pan off the stove with a temperature of 150°C. Assume that the pan is placed on an insulated pad and that a negligible amount of water boils off. What is the temperature when the water and pan reach thermal equilibrium a short time later?
Strategy
The pan is placed on an insulated pad so that little heat transfer occurs with the surroundings. Originally the pan and water are not in thermal equilibrium: the pan is at a higher temperature than the water. Heat transfer then restores thermal equilibrium once the water and pan are in contact. Because heat transfer between the pan and water takes place rapidly, the mass of evaporated water is negligible and the magnitude of the heat lost by the pan is equal to the heat gained by the water. The exchange of heat stops once a thermal equilibrium between the pan and the water is achieved. The heat exchange can be written as $\mid Q_{\text{hot}} \mid = Q_{\text{cold}}$.
Solution

1. Use the equation for heat transfer $Q = \text{mc}\Delta T$ to express the heat lost by the aluminum pan in terms of the mass of the pan, the specific heat of aluminum, the initial temperature of the pan, and the final temperature:
   Equation:

$$Q_{\text{hot}} = m_{\text{Al}}c_{\text{Al}}(T_{\text{f}} - 150^{\circ}\text{C}).$$

2. Express the heat gained by the water in terms of the mass of the water, the specific heat of water, the initial temperature of the water and the final temperature:
   Equation:

$$Q_{\text{cold}} = m_{\text{W}} c_{\text{W}} (T_{\text{f}} - 20.0^{\text{o}}\text{C}).$$

3. Note that $Q_{\text{hot}} < 0$ and $Q_{\text{cold}} > 0$ and that they must sum to zero because the heat lost by the hot pan must be the same as the heat gained by the cold water:
   Equation:

$$\begin{aligned} Q_{\text{cold}} + Q_{\text{hot}} &= 0, \\ Q_{\text{cold}} &= -\text{Q}_{\text{hot}}, \\ m_{\text{W}} c_{\text{W}} (T_{\text{f}} - 20.0^{\text{o}}\text{C}) &= -\text{m}_{\text{Al}} c_{\text{Al}} (T_{\text{f}} - 150^{\text{o}}\text{C}.) \end{aligned}$$

4. This an equation for the unknown final temperature, T_{f}
5. Bring all terms involving T_{f} on the left hand side and all other terms on the right hand side. Solve for T_{f},
   Equation:

$$T_{\text{f}} = \frac{m_{\text{Al}} c_{\text{Al}} (150^{\text{o}}\text{C}) + m_{\text{W}} c_{\text{W}} (20.0^{\text{o}}\text{C})}{m_{\text{Al}} c_{\text{Al}} + m_{\text{W}} c_{\text{W}}},$$

   and insert the numerical values:
   Equation:

$$\begin{aligned} T_{\text{f}} &= \frac{(0.500\ \text{kg})(900\ \text{J/kg}^{\text{o}}\text{C})(150^{\text{o}}\text{C}) + (0.250\ \text{kg})(4186\ \text{J/kg}^{\text{o}}\text{C})(20.0^{\text{o}}\text{C})}{(0.500\ \text{kg})(900\ \text{J/kg}^{\text{o}}\text{C}) + (0.250\ \text{kg})(4186\ \text{J/kg}^{\text{o}}\text{C})} \\ &= \frac{88430\ \text{J}}{1496.5\ \text{J/}^{\text{o}}\text{C}} \\ &= 59.1^{\text{o}}\text{C}. \end{aligned}$$

Discussion

This is a typical *calorimetry* problem—two bodies at different temperatures are brought in contact with each other and exchange heat until a common temperature is reached. Why is the final temperature so much closer to 20.0^{o}C than 150^{o}C? The reason is that water has a greater specific heat than most common substances and thus undergoes a small temperature change for a given heat transfer. A large body of water, such as a lake, requires a large amount of heat to increase its temperature appreciably. This explains why the temperature of a lake stays relatively constant during a day even when the temperature change of the air is large. However, the water temperature does change over longer times (e.g., summer to winter).

Note:
Take-Home Experiment: Temperature Change of Land and Water
What heats faster, land or water?
To study differences in heat capacity:

- Place equal masses of dry sand (or soil) and water at the same temperature into two small jars. (The average density of soil or sand is about 1.6 times that of water, so you can achieve approximately equal masses by using 50% more water by volume.)
- Heat both (using an oven or a heat lamp) for the same amount of time.
- Record the final temperature of the two masses.
- Now bring both jars to the same temperature by heating for a longer period of time.
- Remove the jars from the heat source and measure their temperature every 5 minutes for about 30 minutes.

Which sample cools off the fastest? This activity replicates the phenomena responsible for land breezes and sea breezes.

Exercise:
Check Your Understanding

Problem:

If 25 kJ is necessary to raise the temperature of a block from 25^{o}C to 30^{o}C, how much heat is necessary to heat the block from 45^{o}C to 50^{o}C?

Solution:

The heat transfer depends only on the temperature difference. Since the temperature differences are the same in both cases, the same 25 kJ is necessary in the second case.

Summary

- The transfer of heat Q that leads to a change ΔT in the temperature of a body with mass m is $Q=\text{mc}\Delta T$, where c is the specific heat of the material. This relationship can also be considered as the definition of specific heat.

Conceptual Questions

Exercise:

Problem:

What three factors affect the heat transfer that is necessary to change an object’s temperature?

Exercise:

Problem:

The brakes in a car increase in temperature by ΔT when bringing the car to rest from a speed v. How much greater would ΔT be if the car initially had twice the speed? You may assume the car to stop sufficiently fast so that no heat transfers out of the brakes.

Problems & Exercises

Exercise:

Problem:

On a hot day, the temperature of an 80,000-L swimming pool increases by $1.50^{\mathrm{o}}\mathrm{C}$. What is the net heat transfer during this heating? Ignore any complications, such as loss of water by evaporation.

Solution:
Equation:

$$5.02 \times 10^{8}\ \mathrm{J}$$

Exercise:

Problem: Show that $1\ \mathrm{cal/g}\cdot{}^{\mathrm{o}}\mathrm{C} = 1\ \mathrm{kcal/kg}\cdot{}^{\mathrm{o}}\mathrm{C}$.

Exercise:

Problem:

To sterilize a 50.0-g glass baby bottle, we must raise its temperature from $22.0^{\mathrm{o}}\mathrm{C}$ to $95.0^{\mathrm{o}}\mathrm{C}$. How much heat transfer is required?

Solution:
Equation:

$$3.07 \times 10^{3}\ \mathrm{J}$$

Exercise:

Problem:

The same heat transfer into identical masses of different substances produces different temperature changes. Calculate the final temperature when 1.00 kcal of heat transfers into 1.00 kg of the following, originally at 20.0°C: (a) water; (b) concrete; (c) steel; and (d) mercury.

Exercise:

Problem:

Rubbing your hands together warms them by converting work into thermal energy. If a woman rubs her hands back and forth for a total of 20 rubs, at a distance of 7.50 cm per rub, and with an average frictional force of 40.0 N, what is the temperature increase? The mass of tissues warmed is only 0.100 kg, mostly in the palms and fingers.

Solution:
Equation:

$$0.171^\circ\text{C}$$

Exercise:

Problem:

A 0.250-kg block of a pure material is heated from 20.0°C to 65.0°C by the addition of 4.35 kJ of energy. Calculate its specific heat and identify the substance of which it is most likely composed.

Exercise:

Problem:

Suppose identical amounts of heat transfer into different masses of copper and water, causing identical changes in temperature. What is the ratio of the mass of copper to water?

Solution:

10.8

Exercise:

Problem:

(a) The number of kilocalories in food is determined by calorimetry techniques in which the food is burned and the amount of heat transfer is measured. How many kilocalories per gram are there in a 5.00-g peanut if the energy from burning it is transferred to 0.500 kg of water held in a 0.100-kg aluminum cup, causing a 54.9^{o}C temperature increase? (b) Compare your answer to labeling information found on a package of peanuts and comment on whether the values are consistent.

Exercise:

Problem:

Following vigorous exercise, the body temperature of an 80.0-kg person is 40.0^{o}C. At what rate in watts must the person transfer thermal energy to reduce the the body temperature to 37.0^{o}C in 30.0 min, assuming the body continues to produce energy at the rate of 150 W? $(1\text{ watt} = 1\text{ joule/second or }1\text{ W} = 1\text{ J/s })$.

Solution:

617 W

Exercise:

Problem:

Even when shut down after a period of normal use, a large commercial nuclear reactor transfers thermal energy at the rate of 150 MW by the radioactive decay of fission products. This heat transfer causes a rapid increase in temperature if the cooling system fails $(1\text{ watt} = 1\text{ joule/second or }1\text{ W} = 1\text{ J/s and }1\text{ MW} = 1\text{ megawatt})$. (a) Calculate the rate of temperature increase in degrees Celsius per second $(^{\text{o}}\text{C/s})$ if the mass of the reactor core is $1.60 \times 10^{5}\text{ kg}$ and it has an average specific heat of $0.3349\text{ kJ/kg}^{\text{o}} \cdot \text{C}$. (b) How long would it take to obtain a temperature increase of 2000^{o}C, which could cause some metals holding the radioactive materials to melt? (The initial rate of temperature increase would be greater than that calculated here because the heat transfer is concentrated in a smaller mass. Later, however, the temperature increase would slow down because the 5×10^{5}-kg steel containment vessel would also begin to heat up.)

Radioactive spent-fuel pool at a nuclear power plant. Spent fuel stays hot for a long time. (credit: U.S. Department of Energy)

Glossary

specific heat
: the amount of heat necessary to change the temperature of 1.00 kg of a substance by 1.00 ºC

Phase Change and Latent Heat

- Examine heat transfer.
- Calculate final temperature from heat transfer.

So far we have discussed temperature change due to heat transfer. No temperature change occurs from heat transfer if ice melts and becomes liquid water (i.e., during a phase change). For example, consider water dripping from icicles melting on a roof warmed by the Sun. Conversely, water freezes in an ice tray cooled by lower-temperature surroundings.

Heat from the air transfers to the ice causing it to melt. (credit: Mike Brand)

Energy is required to melt a solid because the cohesive bonds between the molecules in the solid must be broken apart such that, in the liquid, the molecules can move around at comparable kinetic energies; thus, there is no rise in temperature. Similarly, energy is needed to vaporize a liquid, because molecules in a liquid interact with each other via attractive forces. There is no temperature change until a phase change is complete. The temperature of a cup of soda initially at 0ºC stays at 0ºC until all the ice has melted. Conversely, energy is released during freezing and condensation, usually in the form of thermal energy. Work is done by cohesive forces when molecules are brought together. The corresponding energy must be given off (dissipated) to allow them to stay together [link].

The energy involved in a phase change depends on two major factors: the number and strength of bonds or force pairs. The number of bonds is proportional to the number of molecules and thus to the mass of the sample. The strength of forces depends on the type of molecules. The heat Q required to change the phase of a sample of mass m is given by

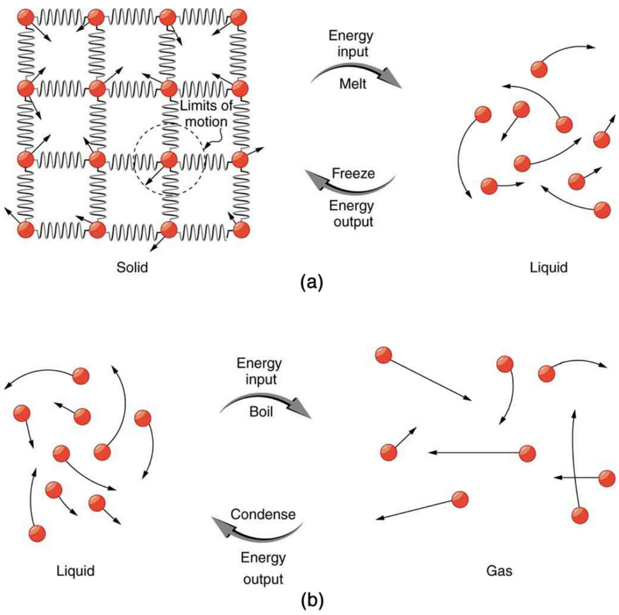
Equation:

$$Q=\mathrm{mL_f}\ (\mathrm{melting/freezing}),$$

Equation:

$$Q=\mathrm{mL_v}\ (\mathrm{vaporization/condensation}),$$

where the latent heat of fusion, L_f, and latent heat of vaporization, L_v, are material constants that are determined experimentally. See ([link]).



(a) Energy is required to partially overcome the attractive forces between molecules in a solid to form a liquid. That same energy must be removed for freezing to take place. (b) Molecules are separated by large distances when going from liquid to vapor, requiring significant energy to overcome molecular attraction. The same energy must be removed for condensation to take place. There is no temperature change until a phase change is complete.

Latent heat is measured in units of J/kg. Both L_f and L_v depend on the substance, particularly on the strength of its molecular forces as noted earlier. L_f and L_v are collectively called **latent heat coefficients**. They are *latent*, or hidden, because in phase changes, energy enters or leaves a system without causing a temperature change in the system; so, in effect, the energy is hidden. [link] lists representative values of L_f and L_v, together with melting and boiling points.

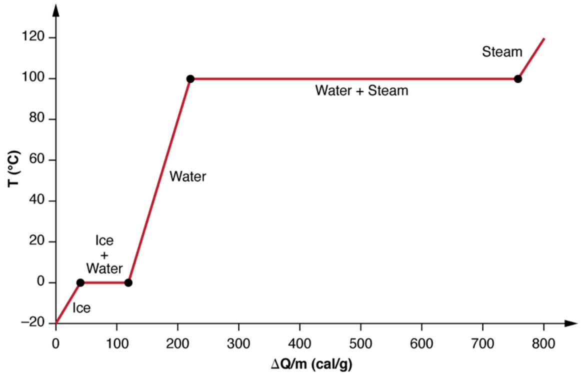
The table shows that significant amounts of energy are involved in phase changes. Let us look, for example, at how much energy is needed to melt a kilogram of ice at 0°C to produce a kilogram of water at $0\,^\circ\text{C}$. Using the equation for a change in temperature and the value for water from [link], we find that $Q = mL_\text{f} = (1.0\text{ kg})(334\text{ kJ/kg}) = 334\text{ kJ}$ is the energy to melt a kilogram of ice. This is a lot of energy as it represents the same amount of energy needed to raise the temperature of 1 kg of liquid water from 0°C to 79.8°C. Even more energy is required to vaporize water; it would take 2256 kJ to change 1 kg of liquid water at the normal boiling point (100°C at atmospheric pressure) to steam (water vapor). This example shows that the energy for a phase change is enormous compared to energy associated with temperature changes without a phase change.

		L_f			L_v	
Substance	**Melting point (ºC)**	**kJ/kg**	**kcal/kg**	**Boiling point (ºC)**	**kJ/kg**	**kcal/kg**
Helium	−269.7	5.23	1.25	−268.9	20.9	4.99
Hydrogen	−259.3	58.6	14.0	−252.9	452	108
Nitrogen	−210.0	25.5	6.09	−195.8	201	48.0
Oxygen	−218.8	13.8	3.30	−183.0	213	50.9
Ethanol	−114	104	24.9	78.3	854	204
Ammonia	−75		108	−33.4	1370	327
Mercury	−38.9	11.8	2.82	357	272	65.0
Water	0.00	334	79.8	100.0	2256[footnote] At 37.0ºC (body temperature), the heat of vaporization L_v for water is 2430 kJ/kg or 580 kcal/kg	539[footnote] At 37.0ºC (body temperature), the heat of vaporization L_v for water is 2430 kJ/kg or 580 kcal/kg
Sulfur	119	38.1	9.10	444.6	326	77.9
Lead	327	24.5	5.85	1750	871	208
Antimony	631	165	39.4	1440	561	134
Aluminum	660	380	90	2450	11400	2720
Silver	961	88.3	21.1	2193	2336	558
Gold	1063	64.5	15.4	2660	1578	377
Copper	1083	134	32.0	2595	5069	1211
Uranium	1133	84	20	3900	1900	454
Tungsten	3410	184	44	5900	4810	1150

Heats of Fusion and Vaporization [footnote]
Values quoted at the normal melting and boiling temperatures at standard atmospheric pressure (1 atm).

Phase changes can have a tremendous stabilizing effect even on temperatures that are not near the melting and boiling points, because evaporation and condensation (conversion of a gas into a liquid state) occur even at temperatures below the boiling point. Take, for example, the fact that air temperatures in humid climates rarely go above 35.0°C, which is because most heat transfer goes into evaporating water into the air. Similarly, temperatures in humid weather rarely fall below the dew point because enormous heat is released when water vapor condenses.

We examine the effects of phase change more precisely by considering adding heat into a sample of ice at -20°C ([link]). The temperature of the ice rises linearly, absorbing heat at a constant rate of $0.50\ \text{cal/g}\cdot^{\circ}\text{C}$ until it reaches 0°C. Once at this temperature, the ice begins to melt until all the ice has melted, absorbing 79.8 cal/g of heat. The temperature remains constant at 0°C during this phase change. Once all the ice has melted, the temperature of the liquid water rises, absorbing heat at a new constant rate of $1.00\ \text{cal/g}\cdot^{\circ}\text{C}$. At 100°C, the water begins to boil and the temperature again remains constant while the water absorbs 539 cal/g of heat during this phase change. When all the liquid has become steam vapor, the temperature rises again, absorbing heat at a rate of $0.482\ \text{cal/g}\cdot^{\circ}\text{C}$.



A graph of temperature versus energy added. The system is constructed so that no vapor evaporates while ice warms to become liquid water, and so that, when vaporization occurs, the vapor remains in of the system. The long stretches of constant temperature values at 0°C and 100°C reflect the large latent heat of melting and vaporization, respectively.

Water can evaporate at temperatures below the boiling point. More energy is required than at the boiling point, because the kinetic energy of water molecules at temperatures below 100°C is less than that at 100°C, hence less energy is available from random thermal motions. Take, for example, the fact that, at body temperature, perspiration from the skin requires a heat input of 2428 kJ/kg, which is about 10 percent higher than the latent heat of vaporization at 100°C. This heat comes from the skin, and thus provides an effective cooling mechanism in hot weather. High humidity inhibits evaporation, so that body temperature might rise, leaving unevaporated sweat on your brow.

Example:
Calculate Final Temperature from Phase Change: Cooling Soda with Ice Cubes

Three ice cubes are used to chill a soda at $20^{\rm o}\text{C}$ with mass $m_{\text{soda}} = 0.25\ \text{kg}$. The ice is at $0^{\rm o}\text{C}$ and each ice cube has a mass of 6.0 g. Assume that the soda is kept in a foam container so that heat loss can be ignored. Assume the soda has the same heat capacity as water. Find the final temperature when all ice has melted.

Strategy

The ice cubes are at the melting temperature of $0^{\rm o}\text{C}$. Heat is transferred from the soda to the ice for melting. Melting of ice occurs in two steps: first the phase change occurs and solid (ice) transforms into liquid water at the melting temperature, then the temperature of this water rises. Melting yields water at $0^{\rm o}\text{C}$, so more heat is transferred from the soda to this water until the water plus soda system reaches thermal equilibrium,

Equation:

$$Q_{\text{ice}} = -Q_{\text{soda}}.$$

The heat transferred to the ice is $Q_{\text{ice}} = m_{\text{ice}}L_{\text{f}} + m_{\text{ice}}c_{\text{W}}(T_{\text{f}} - 0^{\rm o}\text{C})$. The heat given off by the soda is $Q_{\text{soda}} = m_{\text{soda}}c_{\text{W}}(T_{\text{f}} - 20^{\rm o}\text{C})$. Since no heat is lost, $Q_{\text{ice}} = -Q_{\text{soda}}$, so that

Equation:

$$m_{\text{ice}}L_{\text{f}} + m_{\text{ice}}c_{\text{W}}(T_{\text{f}} - 0^{\rm o}\text{C}) = -m_{\text{soda}}c_{\text{W}}(T_{\text{f}} - 20^{\rm o}\text{C}).$$

Bring all terms involving T_{f} on the left-hand-side and all other terms on the right-hand-side. Solve for the unknown quantity T_{f}:

Equation:

$$T_{\text{f}} = \frac{m_{\text{soda}}c_{\text{W}}(20^{\rm o}\text{C}) - m_{\text{ice}}L_{\text{f}}}{(m_{\text{soda}} + m_{\text{ice}})c_{\text{W}}}.$$

Solution

1. Identify the known quantities. The mass of ice is $m_{\text{ice}} = 3{\times}6.0\ \text{g} = 0.018\ \text{kg}$ and the mass of soda is $m_{\text{soda}} = 0.25\ \text{kg}$.
2. Calculate the terms in the numerator:

   Equation:

   $$m_{\text{soda}}c_{\text{W}}(20^{\rm o}\text{C}) = (0.25\ \text{kg})(4186\ \text{J/kg}\cdot^{\rm o}\text{C})(20^{\rm o}\text{C}) = 20{,}930\ \text{J}$$

   and

   Equation:

   $$m_{\text{ice}}L_{\text{f}} = (0.018\ \text{kg})(334{,}000\ \text{J/kg}){=}6012\ \text{J}.$$
3. Calculate the denominator:

   Equation:

   $$(m_{\text{soda}} + m_{\text{ice}})c_{\text{W}} = (0.25\ \text{kg} + 0.018\ \text{kg})(4186\ \text{K/(kg}\cdot^{\rm o}\text{C}){=}1122\ \text{J/}^{\rm o}\text{C}.$$
4. Calculate the final temperature:

   Equation:

   $$T_{\text{f}} = \frac{20{,}930\ \text{J} - 6012\ \text{J}}{1122\ \text{J/}^{\rm o}\text{C}} = 13^{\rm o}\text{C}.$$

Discussion
This example illustrates the enormous energies involved during a phase change. The mass of ice is about 7 percent the mass of water but leads to a noticeable change in the temperature of soda. Although we assumed that the ice was at the freezing temperature, this is incorrect: the typical temperature is $-6^{\circ}\mathrm{C}$. However, this correction gives a final temperature that is essentially identical to the result we found. Can you explain why?

We have seen that vaporization requires heat transfer to a liquid from the surroundings, so that energy is released by the surroundings. Condensation is the reverse process, increasing the temperature of the surroundings. This increase may seem surprising, since we associate condensation with cold objects—the glass in the figure, for example. However, energy must be removed from the condensing molecules to make a vapor condense. The energy is exactly the same as that required to make the phase change in the other direction, from liquid to vapor, and so it can be calculated from $Q = \mathrm{mL_v}$.

Condensation forms on this glass of iced tea because the temperature of the nearby air is reduced to below the dew point. The rate at which water molecules join together exceeds the rate at which they separate, and so water condenses. Energy is released when the water condenses, speeding the melting of the ice in the glass. (credit: Jenny Downing)

Note:
Real-World Application
Energy is also released when a liquid freezes. This phenomenon is used by fruit growers in Florida to protect oranges when the temperature is close to the freezing point $(0^{\circ}\mathrm{C})$. Growers spray water on the

plants in orchards so that the water freezes and heat is released to the growing oranges on the trees. This prevents the temperature inside the orange from dropping below freezing, which would damage the fruit.

The ice on these trees released large amounts of energy when it froze, helping to prevent the temperature of the trees from dropping below 0ºC. Water is intentionally sprayed on orchards to help prevent hard frosts. (credit: Hermann Hammer)

Sublimation is the transition from solid to vapor phase. You may have noticed that snow can disappear into thin air without a trace of liquid water, or the disappearance of ice cubes in a freezer. The reverse is also true: Frost can form on very cold windows without going through the liquid stage. A popular effect is the making of “smoke” from dry ice, which is solid carbon dioxide. Sublimation occurs because the equilibrium vapor pressure of solids is not zero. Certain air fresheners use the sublimation of a solid to inject a perfume into the room. Moth balls are a slightly toxic example of a phenol (an organic compound) that sublimates, while some solids, such as osmium tetroxide, are so toxic that they must be kept in sealed containers to prevent human exposure to their sublimation-produced vapors.

(a)

(b)

Direct transitions between solid and

vapor are common, sometimes useful, and even beautiful. (a) Dry ice sublimates directly to carbon dioxide gas. The visible vapor is made of water droplets. (credit: Windell Oskay) (b) Frost forms patterns on a very cold window, an example of a solid formed directly from a vapor. (credit: Liz West)

All phase transitions involve heat. In the case of direct solid-vapor transitions, the energy required is given by the equation $Q = \text{mL}_\text{s}$, where L_s is the **heat of sublimation**, which is the energy required to change 1.00 kg of a substance from the solid phase to the vapor phase. L_s is analogous to L_f and L_v, and its value depends on the substance. Sublimation requires energy input, so that dry ice is an effective coolant, whereas the reverse process (i.e., frosting) releases energy. The amount of energy required for sublimation is of the same order of magnitude as that for other phase transitions.

The material presented in this section and the preceding section allows us to calculate any number of effects related to temperature and phase change. In each case, it is necessary to identify which temperature and phase changes are taking place and then to apply the appropriate equation. Keep in mind that heat transfer and work can cause both temperature and phase changes.

Problem-Solving Strategies for the Effects of Heat Transfer

1. *Examine the situation to determine that there is a change in the temperature or phase. Is there heat transfer into or out of the system?* When the presence or absence of a phase change is not obvious, you may wish to first solve the problem as if there were no phase changes, and examine the temperature change obtained. If it is sufficient to take you past a boiling or melting point, you should then go back and do the problem in steps—temperature change, phase change, subsequent temperature change, and so on.
2. *Identify and list all objects that change temperature and phase.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns).* A written list is useful.
4. *Make a list of what is given or what can be inferred from the problem as stated (identify the knowns).*
5. *Solve the appropriate equation for the quantity to be determined (the unknown).* If there is a temperature change, the transferred heat depends on the specific heat (see [link]) whereas, for a phase change, the transferred heat depends on the latent heat. See [link].
6. *Substitute the knowns along with their units into the appropriate equation and obtain numerical solutions complete with units.* You will need to do this in steps if there is more than one stage to the process (such as a temperature change followed by a phase change).

7. *Check the answer to see if it is reasonable: Does it make sense?* As an example, be certain that the temperature change does not also cause a phase change that you have not taken into account.

Exercise:
Check Your Understanding

Problem:

Why does snow remain on mountain slopes even when daytime temperatures are higher than the freezing temperature?

Solution:

Snow is formed from ice crystals and thus is the solid phase of water. Because enormous heat is necessary for phase changes, it takes a certain amount of time for this heat to be accumulated from the air, even if the air is above 0°C. The warmer the air is, the faster this heat exchange occurs and the faster the snow melts.

Summary

- Most substances can exist either in solid, liquid, and gas forms, which are referred to as "phases."
- Phase changes occur at fixed temperatures for a given substance at a given pressure, and these temperatures are called boiling and freezing (or melting) points.
- During phase changes, heat absorbed or released is given by:
  Equation:

$$Q = \text{mL},$$

  where L is the latent heat coefficient.

Conceptual Questions

Exercise:

Problem:

Heat transfer can cause temperature and phase changes. What else can cause these changes?

Exercise:

Problem:

How does the latent heat of fusion of water help slow the decrease of air temperatures, perhaps preventing temperatures from falling significantly below 0°C, in the vicinity of large bodies of water?

Exercise:

Problem: What is the temperature of ice right after it is formed by freezing water?

Exercise:

Problem:

If you place 0°C ice into 0°C water in an insulated container, what will happen? Will some ice melt, will more water freeze, or will neither take place?

Exercise:

Problem:

What effect does condensation on a glass of ice water have on the rate at which the ice melts? Will the condensation speed up the melting process or slow it down?

Exercise:

Problem:

In very humid climates where there are numerous bodies of water, such as in Florida, it is unusual for temperatures to rise above about $35^\circ\text{C}(95^\circ\text{F})$. In deserts, however, temperatures can rise far above this. Explain how the evaporation of water helps limit high temperatures in humid climates.

Exercise:

Problem:

In winters, it is often warmer in San Francisco than in nearby Sacramento, 150 km inland. In summers, it is nearly always hotter in Sacramento. Explain how the bodies of water surrounding San Francisco moderate its extreme temperatures.

Exercise:

Problem:

Putting a lid on a boiling pot greatly reduces the heat transfer necessary to keep it boiling. Explain why.

Exercise:

Problem:

Freeze-dried foods have been dehydrated in a vacuum. During the process, the food freezes and must be heated to facilitate dehydration. Explain both how the vacuum speeds up dehydration and why the food freezes as a result.

Exercise:

Problem:

When still air cools by radiating at night, it is unusual for temperatures to fall below the dew point. Explain why.

Exercise:

Problem:

In a physics classroom demonstration, an instructor inflates a balloon by mouth and then cools it in liquid nitrogen. When cold, the shrunken balloon has a small amount of light blue liquid in it, as well as some snow-like crystals. As it warms up, the liquid boils, and part of the crystals sublimate, with some crystals lingering for awhile and then producing a liquid. Identify the blue liquid and the two solids in the cold balloon. Justify your identifications using data from [link].

Problems & Exercises

Exercise:

Problem:

How much heat transfer (in kilocalories) is required to thaw a 0.450-kg package of frozen vegetables originally at 0°C if their heat of fusion is the same as that of water?

Solution:

35.9 kcal

Exercise:

Problem:

A bag containing 0°C ice is much more effective in absorbing energy than one containing the same amount of 0°C water.

a. How much heat transfer is necessary to raise the temperature of 0.800 kg of water from 0°C to 30.0°C?
b. How much heat transfer is required to first melt 0.800 kg of 0°C ice and then raise its temperature?
c. Explain how your answer supports the contention that the ice is more effective.

Exercise:

Problem:

(a) How much heat transfer is required to raise the temperature of a 0.750-kg aluminum pot containing 2.50 kg of water from 30.0°C to the boiling point and then boil away 0.750 kg of water? (b) How long does this take if the rate of heat transfer is 500 W $1\ \text{watt} = 1\ \text{joule/second}\ \ (1\ \text{W} = 1\ \text{J/s})$?

Solution:

(a) 591 kcal

(b) $4.94 \times 10^3\ \text{s}$

Exercise:

Problem:

The formation of condensation on a glass of ice water causes the ice to melt faster than it would otherwise. If 8.00 g of condensation forms on a glass containing both water and 200 g of ice, how many grams of the ice will melt as a result? Assume no other heat transfer occurs.

Exercise:

Problem:

On a trip, you notice that a 3.50-kg bag of ice lasts an average of one day in your cooler. What is the average power in watts entering the ice if it starts at 0°C and completely melts to 0°C water in exactly one day $1\ \text{watt} = 1\ \text{joule/second}\ \ (1\ \text{W} = 1\ \text{J/s})$?

Solution:

13.5 W

Exercise:

Problem:

On a certain dry sunny day, a swimming pool’s temperature would rise by 1.50°C if not for evaporation. What fraction of the water must evaporate to carry away precisely enough energy to keep the temperature constant?

Exercise:

Problem:

(a) How much heat transfer is necessary to raise the temperature of a 0.200-kg piece of ice from -20.0°C to 130°C, including the energy needed for phase changes?
(b) How much time is required for each stage, assuming a constant 20.0 kJ/s rate of heat transfer?
(c) Make a graph of temperature versus time for this process.

Solution:

(a) 148 kcal

(b) 0.418 s, 3.34 s, 4.19 s, 22.6 s, 0.456 s

Exercise:

Problem:

In 1986, a gargantuan iceberg broke away from the Ross Ice Shelf in Antarctica. It was approximately a rectangle 160 km long, 40.0 km wide, and 250 m thick.

(a) What is the mass of this iceberg, given that the density of ice is $917\ \text{kg/m}^3$?

(b) How much heat transfer (in joules) is needed to melt it?

(c) How many years would it take sunlight alone to melt ice this thick, if the ice absorbs an average of $100\ \text{W/m}^2$, 12.00 h per day?

Exercise:

Problem:

How many grams of coffee must evaporate from 350 g of coffee in a 100-g glass cup to cool the coffee from 95.0°C to 45.0°C? You may assume the coffee has the same thermal properties as water and that the average heat of vaporization is 2340 kJ/kg (560 cal/g). (You may neglect the change in mass of the coffee as it cools, which will give you an answer that is slightly larger than correct.)

Solution:

33.0 g

Exercise:

Problem:

(a) It is difficult to extinguish a fire on a crude oil tanker, because each liter of crude oil releases 2.80×10^7 J of energy when burned. To illustrate this difficulty, calculate the number of liters of water that must be expended to absorb the energy released by burning 1.00 L of crude oil, if the water has its temperature raised from $20.0^\mathrm{o}\mathrm{C}$ to $100^\mathrm{o}\mathrm{C}$, it boils, and the resulting steam is raised to $300^\mathrm{o}\mathrm{C}$. (b) Discuss additional complications caused by the fact that crude oil has a smaller density than water.

Solution:

(a) 9.67 L

(b) Crude oil is less dense than water, so it floats on top of the water, thereby exposing it to the oxygen in the air, which it uses to burn. Also, if the water is under the oil, it is less efficient in absorbing the heat generated by the oil.

Exercise:

Problem:

The energy released from condensation in thunderstorms can be very large. Calculate the energy released into the atmosphere for a small storm of radius 1 km, assuming that 1.0 cm of rain is precipitated uniformly over this area.

Exercise:

Problem: To help prevent frost damage, 4.00 kg of $0^\mathrm{o}\mathrm{C}$ water is sprayed onto a fruit tree.

(a) How much heat transfer occurs as the water freezes?

(b) How much would the temperature of the 200-kg tree decrease if this amount of heat transferred from the tree? Take the specific heat to be $3.35\ \mathrm{kJ/kg} \cdot^\mathrm{o} \mathrm{C}$, and assume that no phase change occurs.

Solution:

a) 319 kcal

b) $2.00^\mathrm{o}\mathrm{C}$

Exercise:

Problem:

A 0.250-kg aluminum bowl holding 0.800 kg of soup at $25.0^\mathrm{o}\mathrm{C}$ is placed in a freezer. What is the final temperature if 377 kJ of energy is transferred from the bowl and soup, assuming the soup's thermal properties are the same as that of water? Explicitly show how you follow the steps in Problem-Solving Strategies for the Effects of Heat Transfer.

Exercise:

Problem:

A 0.0500-kg ice cube at $-30.0^\mathrm{o}\mathrm{C}$ is placed in 0.400 kg of $35.0^\mathrm{o}\mathrm{C}$ water in a very well-insulated container. What is the final temperature?

Solution:

$20.6^{\mathrm{o}}\mathrm{C}$

Exercise:

Problem:

If you pour 0.0100 kg of $20.0^{\mathrm{o}}\mathrm{C}$ water onto a 1.20-kg block of ice (which is initially at $-15.0^{\mathrm{o}}\mathrm{C}$), what is the final temperature? You may assume that the water cools so rapidly that effects of the surroundings are negligible.

Exercise:

Problem:

Indigenous people sometimes cook in watertight baskets by placing hot rocks into water to bring it to a boil. What mass of $500^{\mathrm{o}}\mathrm{C}$ rock must be placed in 4.00 kg of $15.0^{\mathrm{o}}\mathrm{C}$ water to bring its temperature to $100^{\mathrm{o}}\mathrm{C}$, if 0.0250 kg of water escapes as vapor from the initial sizzle? You may neglect the effects of the surroundings and take the average specific heat of the rocks to be that of granite.

Solution:

4.38 kg

Exercise:

Problem:

What would be the final temperature of the pan and water in Calculating the Final Temperature When Heat Is Transferred Between Two Bodies: Pouring Cold Water in a Hot Pan if 0.260 kg of water was placed in the pan and 0.0100 kg of the water evaporated immediately, leaving the remainder to come to a common temperature with the pan?

Exercise:

Problem:

In some countries, liquid nitrogen is used on dairy trucks instead of mechanical refrigerators. A 3.00-hour delivery trip requires 200 L of liquid nitrogen, which has a density of $808\ \mathrm{kg/m^3}$.

(a) Calculate the heat transfer necessary to evaporate this amount of liquid nitrogen and raise its temperature to $3.00^{\mathrm{o}}\mathrm{C}$. (Use c_{p} and assume it is constant over the temperature range.) This value is the amount of cooling the liquid nitrogen supplies.

(b) What is this heat transfer rate in kilowatt-hours?

(c) Compare the amount of cooling obtained from melting an identical mass of $0^{\mathrm{o}}\mathrm{C}$ ice with that from evaporating the liquid nitrogen.

Solution:

(a) $1.57\times 10^{4}\ \mathrm{kcal}$

(b) $18.3\ \mathrm{kW}\cdot\mathrm{h}$

(c) $1.29\times 10^{4}\ \mathrm{kcal}$

Exercise:

Problem:

Some gun fanciers make their own bullets, which involves melting and casting the lead slugs. How much heat transfer is needed to raise the temperature and melt 0.500 kg of lead, starting from 25.0°C ?

Glossary

heat of sublimation

the energy required to change a substance from the solid phase to the vapor phase

latent heat coefficient

a physical constant equal to the amount of heat transferred for every 1 kg of a substance during the change in phase of the substance

sublimation

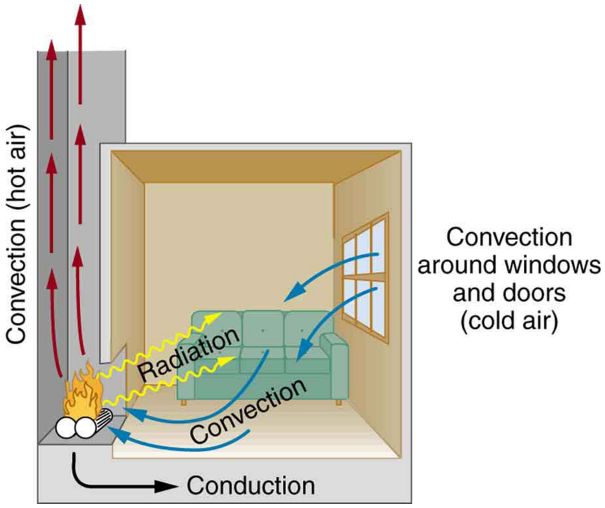
the transition from the solid phase to the vapor phase

Heat Transfer Methods

- Discuss the different methods of heat transfer.

Equally as interesting as the effects of heat transfer on a system are the methods by which this occurs. Whenever there is a temperature difference, heat transfer occurs. Heat transfer may occur rapidly, such as through a cooking pan, or slowly, such as through the walls of a picnic ice chest. We can control rates of heat transfer by choosing materials (such as thick wool clothing for the winter), controlling air movement (such as the use of weather stripping around doors), or by choice of color (such as a white roof to reflect summer sunlight). So many processes involve heat transfer, so that it is hard to imagine a situation where no heat transfer occurs. Yet every process involving heat transfer takes place by only three methods:

1. **Conduction** is heat transfer through stationary matter by physical contact. (The matter is stationary on a macroscopic scale—we know there is thermal motion of the atoms and molecules at any temperature above absolute zero.) Heat transferred between the electric burner of a stove and the bottom of a pan is transferred by conduction.
2. **Convection** is the heat transfer by the macroscopic movement of a fluid. This type of transfer takes place in a forced-air furnace and in weather systems, for example.
3. Heat transfer by **radiation** occurs when microwaves, infrared radiation, visible light, or another form of electromagnetic radiation is emitted or absorbed. An obvious example is the warming of the Earth by the Sun. A less obvious example is thermal radiation from the human body.


Convection (hot air)
Convection around windows and doors (cold air)
Radiation
Convection
Conduction

In a fireplace, heat transfer occurs by all three methods: conduction, convection, and radiation. Radiation is responsible for most of the heat transferred into the room. Heat transfer also occurs through conduction into the room, but at a much slower rate. Heat transfer by convection also occurs through cold air entering the room around windows and hot air leaving the room by rising up the chimney.

We examine these methods in some detail in the three following modules. Each method has unique and interesting characteristics, but all three do have one thing in common: they transfer heat solely because of a temperature difference [link].

Exercise:

Check Your Understanding

Problem:

Name an example from daily life (different from the text) for each mechanism of heat transfer.

Solution:

Conduction: Heat transfers into your hands as you hold a hot cup of coffee.

Convection: Heat transfers as the barista “steams” cold milk to make hot *cocoa*.

Radiation: Reheating a cold cup of coffee in a microwave oven.

Summary

- Heat is transferred by three different methods: conduction, convection, and radiation.

Conceptual Questions

Exercise:

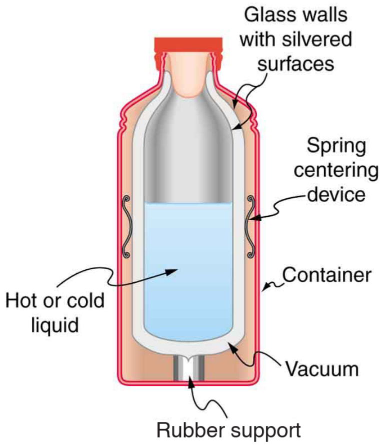
Problem:

What are the main methods of heat transfer from the hot core of Earth to its surface? From Earth’s surface to outer space?

When our bodies get too warm, they respond by sweating and increasing blood circulation to the surface to transfer thermal energy away from the core. What effect will this have on a person in a 40.0°C hot tub?

[link] shows a cut-away drawing of a thermos bottle (also known as a Dewar flask), which is a device designed specifically to slow down all forms of heat transfer. Explain the functions of the various parts, such as the

vacuum, the silvering of the walls, the thin-walled long glass neck, the rubber support, the air layer, and the stopper.


Glass walls with silvered surfaces
Spring centering device
Container
Hot or cold liquid
Vacuum
Rubber support

The construction of a thermos bottle is designed to inhibit all methods of heat transfer.

Glossary

conduction
: heat transfer through stationary matter by physical contact

convection
: heat transfer by the macroscopic movement of fluid

radiation
: heat transfer which occurs when microwaves, infrared radiation, visible light, or other electromagnetic radiation is emitted or absorbed

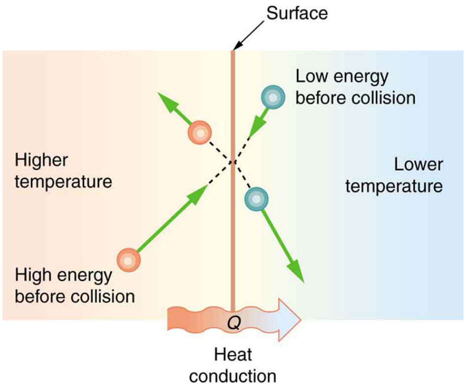
Conduction

- Calculate thermal conductivity.
- Observe conduction of heat in collisions.
- Study thermal conductivities of common substances.

Insulation is used to limit the conduction of heat from the inside to the outside (in winters) and from the outside to the inside (in summers). (credit: Giles Douglas)

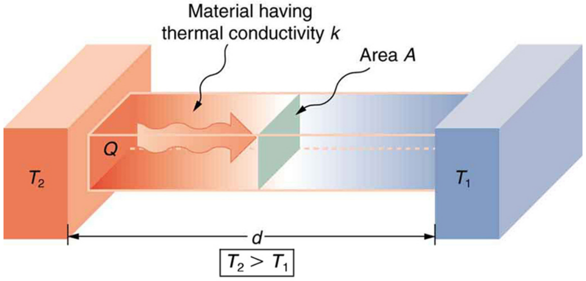
Your feet feel cold as you walk barefoot across the living room carpet in your cold house and then step onto the kitchen tile floor. This result is intriguing, since the carpet and tile floor are both at the same temperature. The different sensation you feel is explained by the different rates of heat transfer: the heat loss during the same time interval is greater for skin in contact with the tiles than with the carpet, so the temperature drop is greater on the tiles.

Some materials conduct thermal energy faster than others. In general, good conductors of electricity (metals like copper, aluminum, gold, and silver) are also good heat conductors, whereas insulators of electricity (wood, plastic, and rubber) are poor heat conductors. [link] shows molecules in two bodies at different temperatures. The (average) kinetic energy of a molecule in the hot body is higher than in the colder body. If two molecules collide, an energy transfer from the molecule with greater kinetic energy to the molecule with less kinetic energy occurs. The cumulative effect from all collisions results in a net flux of heat from the hot body to the colder body. The heat flux thus depends on the temperature difference $\Delta T = T_{\text{hot}} - T_{\text{cold}}$. Therefore, you will get a more severe burn from boiling water than from hot tap water. Conversely, if the temperatures are the same, the net heat transfer rate falls to zero, and equilibrium is achieved. Owing to the fact that the number of collisions increases with increasing area, heat conduction depends on the cross-sectional area. If you touch a cold wall with your palm, your hand cools faster than if you just touch it with your fingertip.


Surface
Low energy before collision
Higher temperature
Lower temperature
High energy before collision
Q
Heat conduction

The molecules in two bodies at different temperatures have different average kinetic energies. Collisions occurring at the contact surface tend to transfer energy from high-temperature regions to low-temperature regions. In this illustration, a molecule in the lower temperature region (right side) has low energy before collision, but its energy increases after colliding with the contact surface. In contrast, a molecule in the higher temperature region (left side) has high energy before collision, but its energy decreases after colliding with the contact surface.

A third factor in the mechanism of conduction is the thickness of the material through which heat transfers. The figure below shows a slab of material with different temperatures on either side. Suppose that T_2 is greater than T_1, so that heat is transferred from left to right. Heat transfer from the left side to the right side is accomplished by a series of molecular collisions. The thicker the material, the more time it takes to transfer the same amount of heat. This model explains why thick clothing is warmer than thin clothing in winters, and why Arctic mammals protect themselves with thick blubber.


Material having thermal conductivity k
Area A
Q
T_2
T_1
d
$T_2 > T_1$

Heat conduction occurs through any material, represented here by a rectangular bar, whether window glass or walrus blubber. The temperature of the material is T_2 on the left and T_1 on the right, where T_2 is greater than T_1.

The rate of heat transfer by conduction is directly proportional to the surface area A, the temperature difference $T_2 - T_1$, and the substance’s conductivity k. The rate of heat transfer is inversely proportional to the thickness d.

Lastly, the heat transfer rate depends on the material properties described by the coefficient of thermal conductivity. All four factors are included in a simple equation that was deduced from and is confirmed by experiments. The **rate of conductive heat transfer** through a slab of material, such as the one in [link], is given by

Equation:

$$\frac{Q}{t} = \frac{\text{kA}(T_2 - T_1)}{d},$$

where Q/t is the rate of heat transfer in watts or kilocalories per second, k is the **thermal conductivity** of the material, A and d are its surface area and thickness, as shown in [link], and $(T_2 - T_1)$ is the temperature difference across the slab. [link] gives representative values of thermal conductivity.

Example:

Calculating Heat Transfer Through Conduction: Conduction Rate Through an Ice Box

A Styrofoam ice box has a total area of $0.950\ \text{m}^2$ and walls with an average thickness of 2.50 cm. The box contains ice, water, and canned beverages at 0°C. The inside of the box is kept cold by melting ice. How much ice melts in one day if the ice box is kept in the trunk of a car at 35.0°C?

Strategy

This question involves both heat for a phase change (melting of ice) and the transfer of heat by conduction. To find the amount of ice melted, we must find the net heat transferred. This value can be obtained by calculating the rate of heat transfer by conduction and multiplying by time.

Solution

1. Identify the knowns.
   Equation:
   $$A = 0.950\ \text{m}^2;\ d = 2.50\ \text{cm} = 0.0250\ \text{m};\ T_1 = 0^\circ\text{C};\ T_2 = 35.0^\circ\text{C},\ t = 1\ \text{day} = 24\ \text{hours} = 86{,}400\ \text{s}.$$
2. Identify the unknowns. We need to solve for the mass of the ice, m. We will also need to solve for the net heat transferred to melt the ice, Q.
3. Determine which equations to use. The rate of heat transfer by conduction is given by
   Equation:
   $$\frac{Q}{t} = \frac{\text{kA}(T_2 - T_1)}{d}.$$
4. The heat is used to melt the ice: $Q = \text{mL}_\text{f}$.
5. Insert the known values:
   Equation:
   $$\frac{Q}{t} = \frac{(0.010\ \text{J/s}\cdot\text{m}\cdot{}^\circ\text{C})\left(0.950\ \text{m}^2\right)(35.0^\circ\text{C} - 0^\circ\text{C})}{0.0250\ \text{m}} = 13.3\ \text{J/s}.$$

6. Multiply the rate of heat transfer by the time (1 day $= 86{,}400$ s):
   Equation:

$$Q = (Q/t)t = (13.3\ \text{J/s})(86{,}400\ \text{s}) = 1.15 \times 10^{6}\ \text{J}.$$

7. Set this equal to the heat transferred to melt the ice: $Q = \text{mL}_{\text{f}}$. Solve for the mass m:
   Equation:

$$m = \frac{Q}{L_{\text{f}}} = \frac{1.15 \times 10^{6}\ \text{J}}{334\ \times 10^{3}\ \text{J/kg}} = 3.44\text{kg}.$$

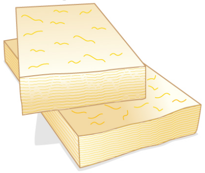
Discussion
The result of 3.44 kg, or about 7.6 lbs, seems about right, based on experience. You might expect to use about a 4 kg (7–10 lb) bag of ice per day. A little extra ice is required if you add any warm food or beverages. Inspecting the conductivities in [link] shows that Styrofoam is a very poor conductor and thus a good insulator. Other good insulators include fiberglass, wool, and goose-down feathers. Like Styrofoam, these all incorporate many small pockets of air, taking advantage of air’s poor thermal conductivity.

Substance	Thermal conductivity k $(\text{J/s}\cdot\text{m}\cdot{}^{\circ}\text{C})$
Silver	420
Copper	390
Gold	318
Aluminum	220
Steel iron	80
Steel (stainless)	14
Ice	2.2
Glass (average)	0.84
Concrete brick	0.84
Water	0.6
Fatty tissue (without blood)	0.2
Asbestos	0.16
Plasterboard	0.16
Wood	0.08–0.16

Substance	Thermal conductivity k $(\mathrm{J/s{\cdot}m{\cdot}^{o}C})$
Snow (dry)	0.10
Cork	0.042
Glass wool	0.042
Wool	0.04
Down feathers	0.025
Air	0.023
Styrofoam	0.010

Thermal Conductivities of Common Substances[footnote]
At temperatures near 0ºC.

A combination of material and thickness is often manipulated to develop good insulators—the smaller the conductivity k and the larger the thickness d, the better. The ratio of d/k will thus be large for a good insulator. The ratio d/k is called the R **factor**. The rate of conductive heat transfer is inversely proportional to R. The larger the value of R, the better the insulation. R factors are most commonly quoted for household insulation, refrigerators, and the like—unfortunately, it is still in non-metric units of $\text{ft}^2\cdot$°F·h/Btu, although the unit usually goes unstated (1 British thermal unit [Btu] is the amount of energy needed to change the temperature of 1.0 lb of water by 1.0 °F). A couple of representative values are an R factor of 11 for 3.5-in-thick fiberglass batts (pieces) of insulation and an R factor of 19 for 6.5-in-thick fiberglass batts. Walls are usually insulated with 3.5-in batts, while ceilings are usually insulated with 6.5-in batts. In cold climates, thicker batts may be used in ceilings and walls.



The fiberglass batt is used for insulation of walls and ceilings to prevent heat transfer between the inside of the building and the outside environment.

Note that in [link], the best thermal conductors—silver, copper, gold, and aluminum—are also the best electrical conductors, again related to the density of free electrons in them. Cooking utensils are typically made from good conductors.

Example:
Calculating the Temperature Difference Maintained by a Heat Transfer: Conduction Through an Aluminum Pan

Water is boiling in an aluminum pan placed on an electrical element on a stovetop. The sauce pan has a bottom that is 0.800 cm thick and 14.0 cm in diameter. The boiling water is evaporating at the rate of 1.00 g/s. What is the temperature difference across (through) the bottom of the pan?

Strategy

Conduction through the aluminum is the primary method of heat transfer here, and so we use the equation for the rate of heat transfer and solve for the temperature difference.

Equation:

$$T_2 - T_1 = \frac{Q}{t}\left(\frac{d}{\mathrm{kA}}\right).$$

Solution

1. Identify the knowns and convert them to the SI units.

   The thickness of the pan, $d = 0.800\ \mathrm{cm} = 8.0 \times 10^{-3}\ \ \mathrm{m}$, the area of the pan, $A = \pi(0.14/2)^2\ \ \mathrm{m}^2 = 1.54 \times 10^{-2}\ \ \mathrm{m}^2$, and the thermal conductivity, $k = 220\ \mathrm{J/s \cdot m \cdot {}^\circ C}$.
2. Calculate the necessary heat of vaporization of 1 g of water:
   Equation:

   $$Q = \mathrm{mL_v} = \left(1.00 \times 10^{-3}\ \mathrm{kg}\right)\left(2256 \times 10^3\ \mathrm{J/kg}\right) = 2256\ \mathrm{J}.$$
3. Calculate the rate of heat transfer given that 1 g of water melts in one second:
   Equation:

   $$Q/t = 2256\ \mathrm{J/s}\ \text{or}\ 2.26\ \mathrm{kW}.$$
4. Insert the knowns into the equation and solve for the temperature difference:
   Equation:

   $$T_2 - T_1 = \frac{Q}{t}\left(\frac{d}{\mathrm{kA}}\right) = (2256\ \mathrm{J/s})\frac{8.00\ \times\ 10^{-3}\mathrm{m}}{(220\ \mathrm{J/s \cdot m \cdot {}^o\ C})\left(1.54 \times 10^{-2}\ \mathrm{m}^2\right)} = 5.33^{\mathrm{o}}\mathrm{C}.$$

Discussion

The value for the heat transfer $Q/t = 2.26\mathrm{kW}$ or $2256\ \mathrm{J/s}$ is typical for an electric stove. This value gives a remarkably small temperature difference between the stove and the pan. Consider that the stove burner is red hot while the inside of the pan is nearly 100ºC because of its contact with boiling water. This contact effectively cools the bottom of the pan in spite of its proximity to the very hot stove burner. Aluminum is such a good conductor that it only takes this small temperature difference to produce a heat transfer of 2.26 kW into the pan.

Conduction is caused by the random motion of atoms and molecules. As such, it is an ineffective mechanism for heat transport over macroscopic distances and short time distances. Take, for example, the temperature on the Earth, which would be unbearably cold during the night and extremely hot during the day if heat transport in the atmosphere was to be only through conduction. In another example, car engines would overheat unless there was a more efficient way to remove excess heat from the pistons.

Exercise:
Check Your Understanding

Problem:

How does the rate of heat transfer by conduction change when all spatial dimensions are doubled?

Solution:

Because area is the product of two spatial dimensions, it increases by a factor of four when each dimension is doubled $\left(A_{\text{final}} = (2d)^2 = 4d^2 = 4A_{\text{initial}}\right)$. The distance, however, simply doubles. Because the temperature difference and the coefficient of thermal conductivity are independent of the spatial dimensions, the rate of heat transfer by conduction increases by a factor of four divided by two, or two:

Equation:

$$\left(\frac{Q}{t}\right)_{\text{final}} = \frac{\mathrm{kA}_{\text{final}}(T_2 - T_1)}{d_{\text{final}}} = \frac{k(4\mathrm{A}_{\text{initial}})(T_2 - T_1)}{2\mathrm{d}_{\text{initial}}} = 2\frac{\mathrm{kA}_{\text{initial}}(T_2 - T_1)}{d_{\text{initial}}} = 2\left(\frac{Q}{t}\right)_{\text{initial}}.$$

Summary

- Heat conduction is the transfer of heat between two objects in direct contact with each other.
- The rate of heat transfer Q/t (energy per unit time) is proportional to the temperature difference $T_2 - T_1$ and the contact area A and inversely proportional to the distance d between the objects:
  Equation:
  $$\frac{Q}{t} = \frac{\mathrm{kA}(T_2 - T_1)}{d}.$$

Conceptual Questions

Exercise:

Problem:

Some electric stoves have a flat ceramic surface with heating elements hidden beneath. A pot placed over a heating element will be heated, while it is safe to touch the surface only a few centimeters away. Why is ceramic, with a conductivity less than that of a metal but greater than that of a good insulator, an ideal choice for the stove top?

Exercise:

Problem:

Loose-fitting white clothing covering most of the body is ideal for desert dwellers, both in the hot Sun and during cold evenings. Explain how such clothing is advantageous during both day and night.

A jellabiya is worn by many men in Egypt. (credit: Zerida)

Problems & Exercises

Exercise:

Problem:

(a) Calculate the rate of heat conduction through house walls that are 13.0 cm thick and that have an average thermal conductivity twice that of glass wool. Assume there are no windows or doors. The surface area of the walls is 120 m^2 and their inside surface is at 18.0°C, while their outside surface is at 5.00°C. (b) How many 1-kW room heaters would be needed to balance the heat transfer due to conduction?

Solution:

(a) $1.01 \times 10^3\text{ W}$

(b) One

Exercise:

Problem:

The rate of heat conduction out of a window on a winter day is rapid enough to chill the air next to it. To see just how rapidly the windows transfer heat by conduction, calculate the rate of conduction in watts through a 3.00-m^2 window that is 0.635 cm thick (1/4 in) if the temperatures of the inner and outer surfaces are 5.00°C and -10.0°C, respectively. This rapid rate will not be maintained—the inner surface will cool, and even result in frost formation.

Exercise:

Problem:

Calculate the rate of heat conduction out of the human body, assuming that the core internal temperature is 37.0°C, the skin temperature is 34.0°C, the thickness of the tissues between averages 1.00 cm, and the surface area is 1.40 m^2.

Solution:

84.0 W

Exercise:

Problem:

Suppose you stand with one foot on ceramic flooring and one foot on a wool carpet, making contact over an area of $80.0\ \text{cm}^2$ with each foot. Both the ceramic and the carpet are 2.00 cm thick and are 10.0^{o}C on their bottom sides. At what rate must heat transfer occur from each foot to keep the top of the ceramic and carpet at 33.0^{o}C?

Exercise:

Problem:

A man consumes 3000 kcal of food in one day, converting most of it to maintain body temperature. If he loses half this energy by evaporating water (through breathing and sweating), how many kilograms of water evaporate?

Solution:

2.59 kg

Exercise:

Problem:

(a) A firewalker runs across a bed of hot coals without sustaining burns. Calculate the heat transferred by conduction into the sole of one foot of a firewalker given that the bottom of the foot is a 3.00-mm-thick callus with a conductivity at the low end of the range for wood and its density is $300\ \text{kg/m}^3$. The area of contact is $25.0\ \text{cm}^2$, the temperature of the coals is 700^{o}C, and the time in contact is 1.00 s.

(b) What temperature increase is produced in the $25.0\ \text{cm}^3$ of tissue affected?

(c) What effect do you think this will have on the tissue, keeping in mind that a callus is made of dead cells?

Exercise:

Problem:

(a) What is the rate of heat conduction through the 3.00-cm-thick fur of a large animal having a 1.40-m^2 surface area? Assume that the animal's skin temperature is 32.0^{o}C, that the air temperature is -5.00^{o}C, and that fur has the same thermal conductivity as air. (b) What food intake will the animal need in one day to replace this heat transfer?

Solution:

(a) 39.7 W

(b) 820 kcal

Exercise:

Problem:

A walrus transfers energy by conduction through its blubber at the rate of 150 W when immersed in -1.00ºC water. The walrus's internal core temperature is 37.0ºC, and it has a surface area of $2.00\ \text{m}^2$. What is the average thickness of its blubber, which has the conductivity of fatty tissues without blood?

Walrus on ice. (credit: Captain Budd Christman, NOAA Corps)

Exercise:

Problem:

Compare the rate of heat conduction through a 13.0-cm-thick wall that has an area of $10.0\ \text{m}^2$ and a thermal conductivity twice that of glass wool with the rate of heat conduction through a window that is 0.750 cm thick and that has an area of $2.00\ \text{m}^2$, assuming the same temperature difference across each.

Solution:

35 to 1, window to wall

Exercise:

Problem:

Suppose a person is covered head to foot by wool clothing with average thickness of 2.00 cm and is transferring energy by conduction through the clothing at the rate of 50.0 W. What is the temperature difference across the clothing, given the surface area is $1.40\ \text{m}^2$?

Exercise:

Problem:

Some stove tops are smooth ceramic for easy cleaning. If the ceramic is 0.600 cm thick and heat conduction occurs through the same area and at the same rate as computed in [link], what is the temperature difference across it? Ceramic has the same thermal conductivity as glass and brick.

Solution:

$1.05\times10^3\ \text{K}$

Exercise:

Problem:

One easy way to reduce heating (and cooling) costs is to add extra insulation in the attic of a house. Suppose the house already had 15 cm of fiberglass insulation in the attic and in all the exterior surfaces. If you added an extra 8.0 cm of fiberglass to the attic, then by what percentage would the heating cost of the house drop? Take the single story house to be of dimensions 10 m by 15 m by 3.0 m. Ignore air infiltration and heat loss through windows and doors.

Exercise:

Problem:

(a) Calculate the rate of heat conduction through a double-paned window that has a 1.50-m^2 area and is made of two panes of 0.800-cm-thick glass separated by a 1.00-cm air gap. The inside surface temperature is 15.0°C, while that on the outside is -10.0°C. (Hint: There are identical temperature drops across the two glass panes. First find these and then the temperature drop across the air gap. This problem ignores the increased heat transfer in the air gap due to convection.)

(b) Calculate the rate of heat conduction through a 1.60-cm-thick window of the same area and with the same temperatures. Compare your answer with that for part (a).

Solution:

(a) 83 W

(b) 24 times that of a double pane window.

Exercise:

Problem:

Many decisions are made on the basis of the payback period: the time it will take through savings to equal the capital cost of an investment. Acceptable payback times depend upon the business or philosophy one has. (For some industries, a payback period is as small as two years.) Suppose you wish to install the extra insulation in [link]. If energy cost $1.00 per million joules and the insulation was $4.00 per square meter, then calculate the simple payback time. Take the average ΔT for the 120 day heating season to be 15.0°C.

Exercise:

Problem:

For the human body, what is the rate of heat transfer by conduction through the body's tissue with the following conditions: the tissue thickness is 3.00 cm, the change in temperature is 2.00°C, and the skin area is $1.50\ \text{m}^2$. How does this compare with the average heat transfer rate to the body resulting from an energy intake of about 2400 kcal per day? (No exercise is included.)

Solution:

20.0 W, 17.2% of 2400 kcal per day

Glossary

R factor	the ratio of thickness to the conductivity of a material
rate of conductive heat transfer	rate of heat transfer from one material to another

thermal conductivity

the property of a material’s ability to conduct heat

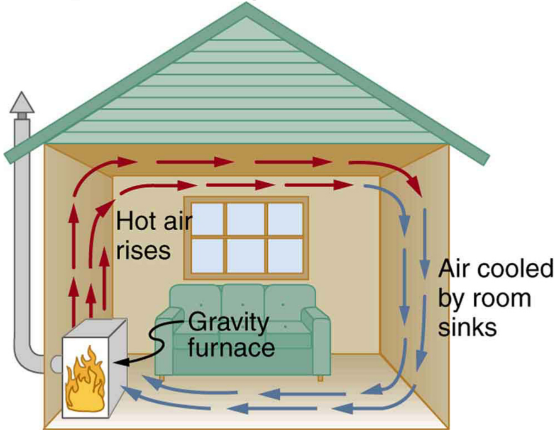
Convection

- Discuss the method of heat transfer by convection.

Convection is driven by large-scale flow of matter. In the case of Earth, the atmospheric circulation is caused by the flow of hot air from the tropics to the poles, and the flow of cold air from the poles toward the tropics. (Note that Earth's rotation causes the observed easterly flow of air in the northern hemisphere). Car engines are kept cool by the flow of water in the cooling system, with the water pump maintaining a flow of cool water to the pistons. The circulatory system is used the body: when the body overheats, the blood vessels in the skin expand (dilate), which increases the blood flow to the skin where it can be cooled by sweating. These vessels become smaller when it is cold outside and larger when it is hot (so more fluid flows, and more energy is transferred).

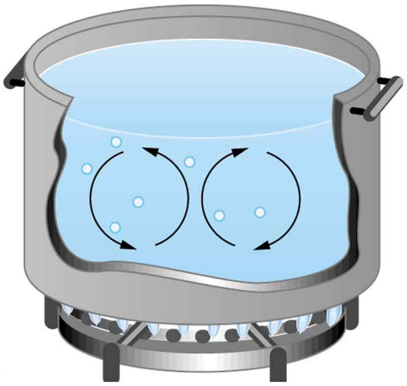
The body also loses a significant fraction of its heat through the breathing process.

While convection is usually more complicated than conduction, we can describe convection and do some straightforward, realistic calculations of its effects. Natural convection is driven by buoyant forces: hot air rises because density decreases as temperature increases. The house in [link] is kept warm in this manner, as is the pot of water on the stove in [link]. Ocean currents and large-scale atmospheric circulation transfer energy from one part of the globe to another. Both are examples of natural convection.



Air heated by the so-called

gravity furnace expands and rises, forming a convective loop that transfers energy to other parts of the room. As the air is cooled at the ceiling and outside walls, it contracts, eventually becoming denser than room air and sinking to the floor. A properly designed heating system using natural convection, like this one, can be quite efficient in uniformly heating a home.



Convection plays an important role in heat transfer inside this pot of water. Once conducted to the inside, heat transfer to other parts of the pot is mostly by convection. The hotter water expands, decreases in density, and rises to transfer heat to other regions of the water, while colder water sinks to the bottom. This process keeps repeating.

Note:

Take-Home Experiment: Convection Rolls in a Heated Pan

Take two small pots of water and use an eye dropper to place a drop of food coloring near the bottom of each. Leave one on a bench top and heat the other over a stovetop. Watch how the color spreads and how long it takes the color to reach the top. Watch how convective loops form.

Example:

Calculating Heat Transfer by Convection: Convection of Air Through the Walls of a House

Most houses are not airtight: air goes in and out around doors and windows, through cracks and crevices, following wiring to switches and outlets, and so on. The air in a typical house is completely replaced in less than an hour. Suppose that a moderately-sized house has inside dimensions $12.0\text{m} \times 18.0\text{m} \times 3.00\text{m}$ high, and that all air is replaced in 30.0 min. Calculate the heat transfer per unit time in watts needed to warm the incoming cold air by 10.0°C, thus replacing the heat transferred by convection alone.

Strategy

Heat is used to raise the temperature of air so that $Q = \text{mc}\Delta T$. The rate of heat transfer is then Q/t, where t is the time for air turnover. We are given that ΔT is 10.0°C, but we must still find values for the mass of air and its specific heat before we can calculate Q. The specific heat of air is a weighted average of the specific heats of nitrogen and oxygen, which gives $c = c_{\text{p}} \cong 1000\ \text{J/kg} \cdot^{\text{o}} \text{C}$ from [link] (note that the specific heat at constant pressure must be used for this process).

Solution

1. Determine the mass of air from its density and the given volume of the house. The density is given from the density ρ and the volume
   Equation:

   $$m = \text{ρV} = \left(1.29\ \text{kg/m}^3\right)(12.0\ \text{m} \times 18.0\ \text{m} \times 3.00\ \text{m}) = 836\ \text{kg}.$$

2. Calculate the heat transferred from the change in air temperature: $Q = \text{mc}\Delta T$ so that
   Equation:

   $$Q = (836\ \text{kg})(1000\ \text{J/kg} \cdot^{\text{o}} \text{C})(10.0^{\text{o}}\text{C}) = 8.36 \times 10^6\ \text{J}.$$

3. Calculate the heat transfer from the heat Q and the turnover time t. Since air is turned over in $t = 0.500\ \text{h} = 1800\ \text{s}$, the heat transferred per unit time is
   Equation:

   $$\frac{Q}{t} = \frac{8.36 \times 10^6\ \text{J}}{1800\ \text{s}} = 4.64\ \text{kW}.$$

Discussion

This rate of heat transfer is equal to the power consumed by about forty-six 100-W light bulbs. Newly constructed homes are designed for a turnover time of 2 hours or more, rather than 30 minutes for the house of this example. Weather stripping, caulking, and improved window seals are commonly employed. More extreme measures are sometimes taken in very cold (or hot) climates to achieve a tight standard of more than 6 hours for one air turnover. Still longer turnover times are unhealthy, because a minimum amount of fresh air is necessary to supply oxygen for breathing and to dilute household pollutants. The term used for the process by which outside air leaks into the house from cracks around windows, doors, and the foundation is called “air infiltration.”

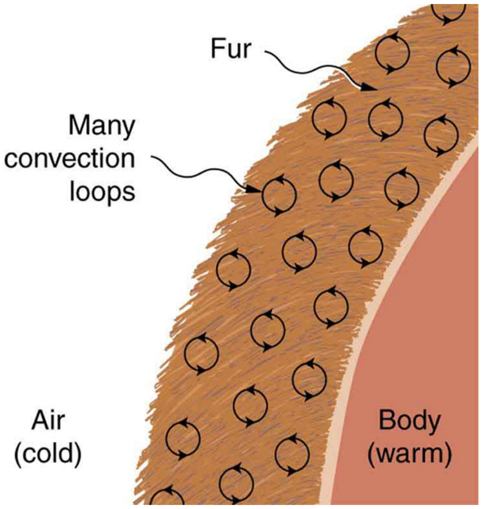
A cold wind is much more chilling than still cold air, because convection combines with conduction in the body to increase the rate at which energy is transferred away from the body. The table below gives approximate wind-chill factors, which are the temperatures of still air that produce the same rate of cooling as air of a given temperature and speed. Wind-chill factors are a dramatic reminder of convection’s ability to transfer heat faster than conduction. For example, a 15.0 m/s wind at 0°C has the chilling equivalent of still air at about -18°C.

Moving air temperature	**Wind speed (m/s)**				
(ºC)	**2**	**5**	**10**	**15**	**20**
5	3	−1	−8	−10	−12
2	0	−7	−12	−16	−18
0	−2	−9	−15	−18	−20

Moving air temperature	**Wind speed (m/s)**				
−5	−7	−15	−22	−26	−29
−10	−12	−21	−29	−34	−36
−20	−23	−34	−44	−50	−52
−40	−44	−59	−73	−82	−84

Wind-Chill Factors

Although air can transfer heat rapidly by convection, it is a poor conductor and thus a good insulator. The amount of available space for airflow determines whether air acts as an insulator or conductor. The space between the inside and outside walls of a house, for example, is about 9 cm (3.5 in)—large enough for convection to work effectively. The addition of wall insulation prevents airflow, so heat loss (or gain) is decreased. Similarly, the gap between the two panes of a double-paned window is about 1 cm, which prevents convection and takes advantage of air’s low conductivity to prevent greater loss. Fur, fiber, and fiberglass also take advantage of the low conductivity of air by trapping it in spaces too small to support convection, as shown in the figure. Fur and feathers are lightweight and thus ideal for the protection of animals.



Fur is filled with air, breaking it up into many small pockets. Convection is very slow here, because the loops are so small. The low conductivity of air makes fur a very good lightweight insulator.

Some interesting phenomena happen *when convection is accompanied by a phase change*. It allows us to cool off by sweating, even if the temperature of the surrounding air exceeds body temperature. Heat from the skin is required for sweat to evaporate from the skin, but without air flow, the air becomes saturated and evaporation stops. Air flow caused by convection replaces the saturated air by dry air and evaporation continues.

Example:

Calculate the Flow of Mass during Convection: Sweat-Heat Transfer away from the Body

The average person produces heat at the rate of about 120 W when at rest. At what rate must water evaporate from the body to get rid of all this energy? (This evaporation might occur when a person is sitting in the shade and surrounding temperatures are the same as skin temperature, eliminating heat transfer by other methods.)

Strategy

Energy is needed for a phase change ($Q = \text{mL}_\text{v}$). Thus, the energy loss per unit time is

Equation:

$$\frac{Q}{t} = \frac{\text{mL}_\text{v}}{t} = 120 \;\text{ W} = 120\text{ J/s}.$$

We divide both sides of the equation by L_v to find that the mass evaporated per unit time is

Equation:

$$\frac{m}{t} = \frac{120\text{ J/s}}{L_\text{v}}.$$

Solution

(1) Insert the value of the latent heat from [link], $L_\text{v} = 2430\text{ kJ/kg} = 2430\text{ J/g}$. This yields

Equation:

$$\frac{m}{t} = \frac{120\text{ J/s}}{2430\text{ J/g}} = 0.0494\text{ g/s} = 2.96\text{ g/min}.$$

Discussion

Evaporating about 3 g/min seems reasonable. This would be about 180 g (about 7 oz) per hour. If the air is very dry, the sweat may evaporate without even being noticed. A significant amount of evaporation also takes place in the lungs and breathing passages.

Another important example of the combination of phase change and convection occurs when water evaporates from the oceans. Heat is removed from the ocean when water evaporates. If the water vapor condenses in liquid droplets as clouds form, heat is released in the atmosphere. Thus, there is an overall transfer of heat from the ocean to the atmosphere. This process is the driving power behind thunderheads, those great cumulus clouds that rise as much as 20.0 km into the stratosphere. Water vapor carried in by convection condenses, releasing tremendous amounts of energy. This energy causes the air to expand and rise, where it is colder. More condensation occurs in these colder regions, which in turn drives the cloud even higher. Such a mechanism is called positive feedback, since the process reinforces and accelerates itself. These systems sometimes produce violent storms, with lightning and hail, and constitute the mechanism driving hurricanes.

Cumulus clouds are caused by water vapor that rises because of convection. The rise of clouds is driven by a positive feedback mechanism. (credit: Mike Love)

Convection accompanied by a phase change releases the energy needed to drive this thunderhead into the stratosphere. (credit: Gerardo García Moretti)

The phase change that occurs when this iceberg melts involves tremendous heat

transfer. (credit:
Dominic Alves)

The movement of icebergs is another example of convection accompanied by a phase change. Suppose an iceberg drifts from Greenland into warmer Atlantic waters. Heat is removed from the warm ocean water when the ice melts and heat is released to the land mass when the iceberg forms on Greenland.

Exercise:

Check Your Understanding

Problem: Explain why using a fan in the summer feels refreshing!

Solution:

Using a fan increases the flow of air: warm air near your body is replaced by cooler air from elsewhere. Convection increases the rate of heat transfer so that moving air “feels” cooler than still air.

Summary

- Convection is heat transfer by the macroscopic movement of mass. Convection can be natural or forced and generally transfers thermal energy faster than conduction. [link] gives wind-chill factors, indicating that moving air has the same chilling effect of much colder stationary air. *Convection that occurs along with a phase change* can transfer energy from cold regions to warm ones.

Conceptual Questions

Exercise:

Problem:

One way to make a fireplace more energy efficient is to have an external air supply for the combustion of its fuel. Another is to have room air circulate around the outside of the fire box and back into the room. Detail the methods of heat transfer involved in each.

Exercise:

Problem:

On cold, clear nights horses will sleep under the cover of large trees. How does this help them keep warm?

Problems & Exercises

Exercise:

Problem:

At what wind speed does -10°C air cause the same chill factor as still air at -29°C?

Solution:

10 m/s

Exercise:

Problem:

At what temperature does still air cause the same chill factor as -5°C air moving at 15 m/s?

Exercise:

Problem:

The “steam” above a freshly made cup of instant coffee is really water vapor droplets condensing after evaporating from the hot coffee. What is the final temperature of 250 g of hot coffee initially at 90.0°C if 2.00 g evaporates from it? The coffee is in a Styrofoam cup, so other methods of heat transfer can be neglected.

Solution:

85.7°C

Exercise:

Problem:

(a) How many kilograms of water must evaporate from a 60.0-kg woman to lower her body temperature by 0.750°C?

(b) Is this a reasonable amount of water to evaporate in the form of perspiration, assuming the relative humidity of the surrounding air is low?

Exercise:

Problem:

On a hot dry day, evaporation from a lake has just enough heat transfer to balance the $1.00\ \text{kW/m}^2$ of incoming heat from the Sun. What mass of water evaporates in 1.00 h from each square meter? Explicitly show how you follow the steps in the Problem-Solving Strategies for the Effects of Heat Transfer.

Solution:

1.48 kg

Exercise:

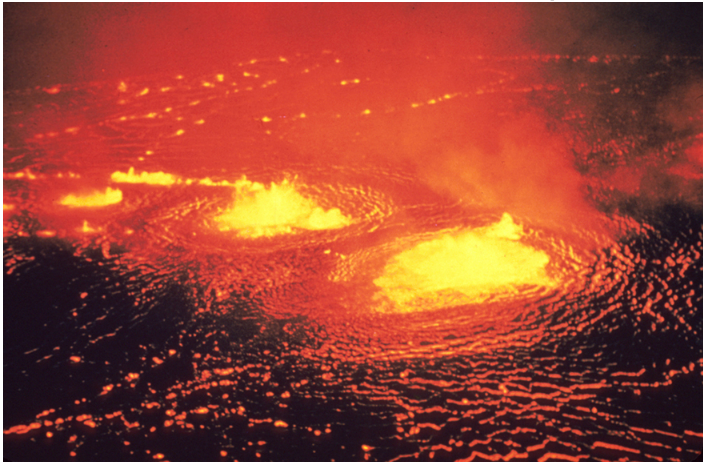
Problem:

One winter day, the climate control system of a large university classroom building malfunctions. As a result, $500\ \text{m}^3$ of excess cold air is brought in each minute. At what rate in kilowatts must heat transfer occur to warm this air by 10.0°C (that is, to bring the air to room temperature)?

Exercise:

Problem:

The Kilauea volcano in Hawaii is the world's most active, disgorging about $5 \times 10^5\ \text{m}^3$ of 1200°C lava per day. What is the rate of heat transfer out of Earth by convection if this lava has a density of $2700\ \text{kg/m}^3$ and eventually cools to 30°C? Assume that the specific heat of lava is the same as that of granite.



Lava flow on Kilauea volcano in Hawaii. (credit: J. P. Eaton, U.S. Geological Survey)

Solution:

$2 \times 10^4\ \text{MW}$

Exercise:

Problem:

During heavy exercise, the body pumps 2.00 L of blood per minute to the surface, where it is cooled by $2.00^{\circ}\mathrm{C}$. What is the rate of heat transfer from this forced convection alone, assuming blood has the same specific heat as water and its density is $1050\ \mathrm{kg/m^3}$?

Exercise:

Problem:

A person inhales and exhales 2.00 L of $37.0^{\circ}\mathrm{C}$ air, evaporating 4.00×10^{-2} g of water from the lungs and breathing passages with each breath.

(a) How much heat transfer occurs due to evaporation in each breath?

(b) What is the rate of heat transfer in watts if the person is breathing at a moderate rate of 18.0 breaths per minute?

(c) If the inhaled air had a temperature of $20.0^{\circ}\mathrm{C}$, what is the rate of heat transfer for warming the air?

(d) Discuss the total rate of heat transfer as it relates to typical metabolic rates. Will this breathing be a major form of heat transfer for this person?

Solution:

(a) 97.2 J

(b) 29.2 W

(c) 9.49 W

(d) The total rate of heat loss would be $29.2\ \mathrm{W} + 9.49\ \mathrm{W} = 38.7\ \mathrm{W}$. While sleeping, our body consumes 83 W of power, while sitting it consumes 120 to 210 W. Therefore, the total rate of heat loss from breathing will not be a major form of heat loss for this person.

Exercise:

Problem:

A glass coffee pot has a circular bottom with a 9.00-cm diameter in contact with a heating element that keeps the coffee warm with a continuous heat transfer rate of 50.0 W

(a) What is the temperature of the bottom of the pot, if it is 3.00 mm thick and the inside temperature is 60.0°C?

(b) If the temperature of the coffee remains constant and all of the heat transfer is removed by evaporation, how many grams per minute evaporate? Take the heat of vaporization to be 2340 kJ/kg.

Radiation

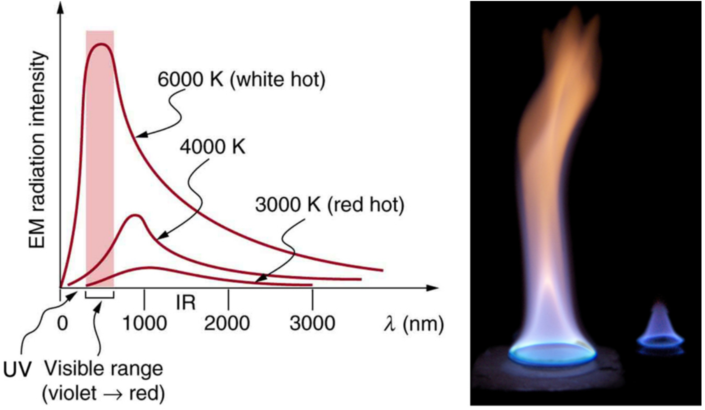
- Discuss heat transfer by radiation.
- Explain the power of different materials.

You can feel the heat transfer from a fire and from the Sun. Similarly, you can sometimes tell that the oven is hot without touching its door or looking inside—it may just warm you as you walk by. The space between the Earth and the Sun is largely empty, without any possibility of heat transfer by convection or conduction. In these examples, heat is transferred by radiation. That is, the hot body emits electromagnetic waves that are absorbed by our skin: no medium is required for electromagnetic waves to propagate. Different names are used for electromagnetic waves of different wavelengths: radio waves, microwaves, infrared **radiation**, visible light, ultraviolet radiation, X-rays, and gamma rays.

Most of the heat transfer from this fire to the observers is through infrared radiation. The visible light, although dramatic, transfers relatively little thermal energy. Convection transfers energy away from the observers as hot air rises, while conduction is negligibly slow here. Skin is very sensitive to infrared radiation, so that you can sense the presence of a fire without looking at it directly. (credit: Daniel X. O’Neil)

The energy of electromagnetic radiation depends on the wavelength (color) and varies over a wide range: a smaller wavelength (or higher frequency) corresponds to a higher energy. Because more heat is radiated at higher temperatures, a temperature change is accompanied by a color change. Take, for example, an electrical element on a stove, which glows from red to orange, while the higher-temperature steel in a blast furnace glows from yellow to white. The radiation you feel is mostly infrared, which corresponds to a lower temperature than that of the electrical element and the steel. The radiated energy depends on its intensity, which is represented in the figure below by the height of the distribution.

Electromagnetic Waves explains more about the electromagnetic spectrum and Introduction to Quantum Physics discusses how the decrease in wavelength corresponds to an increase in energy.

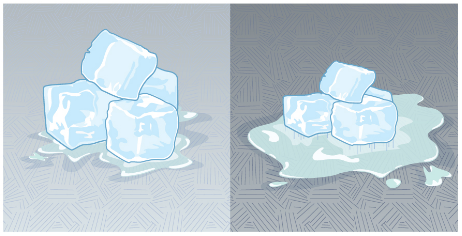


(a)

(b)

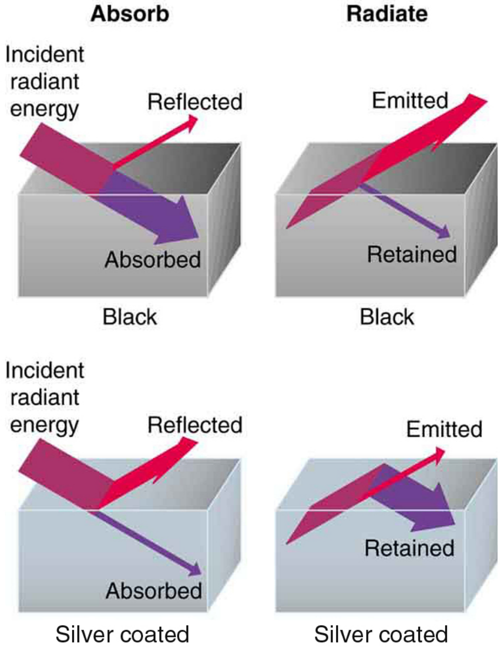
(a) A graph of the spectra of electromagnetic waves emitted from an ideal radiator at three different temperatures. The intensity or rate of radiation emission increases dramatically with temperature, and the spectrum shifts toward the visible and ultraviolet parts of the spectrum. The shaded portion denotes the visible part of the spectrum. It is apparent that the shift toward the ultraviolet with temperature makes the visible appearance shift from red to white to blue as temperature increases. (b) Note the variations in color corresponding to variations in flame temperature. (credit: Tuohirulla)

All objects absorb and emit electromagnetic radiation. The rate of heat transfer by radiation is largely determined by the color of the object. Black is the most effective, and white is the least effective. People living in hot climates generally avoid wearing black clothing, for instance (see [link]). Similarly, black asphalt in a parking lot will be hotter than adjacent gray sidewalk on a summer day, because black absorbs better than gray. The reverse is also true—black radiates better than gray. Thus, on a clear summer night, the asphalt will be colder than the gray sidewalk, because black radiates the energy more rapidly than gray. An *ideal radiator* is the same color as an *ideal absorber*, and captures all the radiation that falls on it. In contrast, white is a poor absorber and is also a poor radiator. A white object reflects all radiation, like a mirror. (A perfect, polished white surface is mirror-like in appearance, and a crushed mirror looks white.)



This illustration shows that the darker pavement is hotter than the lighter pavement (much more of the ice on the right has melted), although both have been in the sunlight for the same time. The thermal conductivities of the pavements are the same.

Gray objects have a uniform ability to absorb all parts of the electromagnetic spectrum. Colored objects behave in similar but more complex ways, which gives them a particular color in the visible range and may make them special in other ranges of the nonvisible spectrum. Take, for example, the strong absorption of infrared radiation by the skin, which allows us to be very sensitive to it.


Absorb
Radiate
Incident radiant energy
Reflected
Absorbed
Black
Emitted
Retained
Black
Incident radiant energy
Reflected
Absorbed
Silver coated
Emitted
Retained
Silver coated

A black object is a good absorber and a good radiator, while a white (or silver) object is a poor absorber and a poor radiator. It is as if radiation from the inside is reflected back into the silver object, whereas radiation from the inside of the black object is "absorbed" when it hits the surface and finds itself on the outside and is strongly emitted.

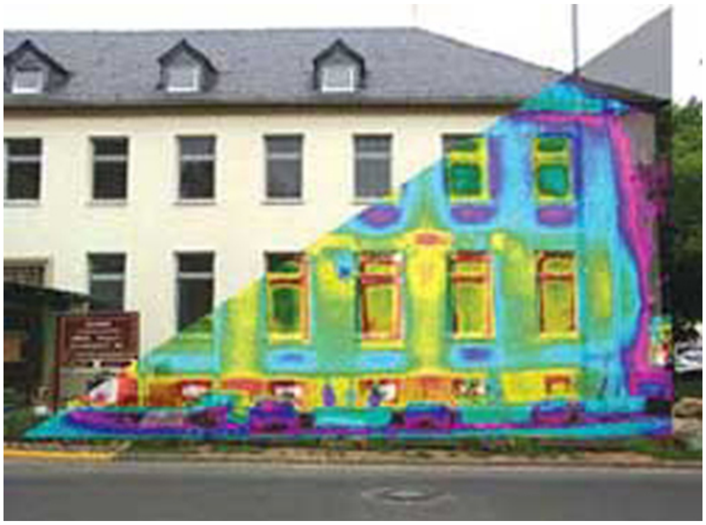
The rate of heat transfer by emitted radiation is determined by the **Stefan-Boltzmann law of radiation**:

Equation:

$$\frac{Q}{t} = \sigma e A T^4,$$

where $\sigma = 5.67 \times 10^{-8}\ \mathrm{J/s \cdot m^2 \cdot K^4}$ is the Stefan-Boltzmann constant, A is the surface area of the object, and T is its absolute temperature in kelvin. The symbol e stands for the **emissivity** of the object, which is a measure of how well it radiates. An ideal jet-black (or black body) radiator has $e = 1$, whereas a perfect reflector has $e = 0$. Real objects fall between these two values. Take, for example, tungsten light bulb filaments which have an e of about 0.5, and carbon black (a material used in printer toner), which has the (greatest known) emissivity of about 0.99.

The radiation rate is directly proportional to the *fourth power* of the absolute temperature—a remarkably strong temperature dependence. Furthermore, the radiated heat is proportional to the surface area of the object. If you knock apart the coals of a fire, there is a noticeable increase in radiation due to an increase in radiating surface area.



A thermograph of part of a building shows temperature variations, indicating where heat transfer to the outside is most severe. Windows are a major region of heat transfer to the outside of homes. (credit: U.S. Army)

Skin is a remarkably good absorber and emitter of infrared radiation, having an emissivity of 0.97 in the infrared spectrum. Thus, we are all nearly (jet) black in the infrared, in spite of the obvious variations in skin color. This high infrared emissivity is why we can so easily feel radiation on our skin. It is also the basis for the use of night scopes used by law enforcement and the military to detect human beings. Even small temperature variations can be detected because of the T^4 dependence. Images, called *thermographs*, can be used medically to detect regions of abnormally high temperature in the body, perhaps indicative of disease. Similar techniques can be used to detect heat leaks in homes [link], optimize performance of blast furnaces, improve comfort levels in work environments, and even remotely map the Earth’s temperature profile.

All objects emit and absorb radiation. The *net* rate of heat transfer by radiation (absorption minus emission) is related to both the temperature of the object and the temperature of its surroundings. Assuming that an object with a temperature T_1 is surrounded by an environment with uniform temperature T_2, the **net rate of heat transfer by radiation** is

Equation:

$$\frac{Q_{\text{net}}}{t} = \sigma e A\left(T_2^4 - T_1^4\right),$$

where e is the emissivity of the object alone. In other words, it does not matter whether the surroundings are white, gray, or black; the balance of radiation into and out of the object depends on how well it emits and absorbs radiation. When $T_2 > T_1$, the quantity Q_{net}/t is positive; that is, the net heat transfer is from hot to cold.

Note:

Take-Home Experiment: Temperature in the Sun

Place a thermometer out in the sunshine and shield it from direct sunlight using an aluminum foil. What is the reading? Now remove the shield, and note what the thermometer reads. Take a handkerchief soaked in nail polish remover, wrap it around the thermometer and place it in the sunshine. What does the thermometer read?

Example:

Calculate the Net Heat Transfer of a Person: Heat Transfer by Radiation

What is the rate of heat transfer by radiation, with an unclothed person standing in a dark room whose ambient temperature is 22.0°C. The person has a normal skin temperature of 33.0°C and a surface area of $1.50\ \text{m}^2$. The emissivity of skin is 0.97 in the infrared, where the radiation takes place.

Strategy

We can solve this by using the equation for the rate of radiative heat transfer.

Solution

Insert the temperatures values $T_2 = 295\text{ K}$ and $T_1 = 306\text{ K}$, so that

Equation:

$$\frac{Q}{t} = \sigma e A\left(T_2^4 - T_1^4\right)$$

Equation:

$$= \left(5.67 \times 10^{-8}\text{ J/s}\cdot\text{ m}^2\cdot\text{ K}^4\right)(0.97)\left(1.50\text{ m}^2\right)\left[(295\text{ K})^4 - (306\text{ K})^4\right]$$

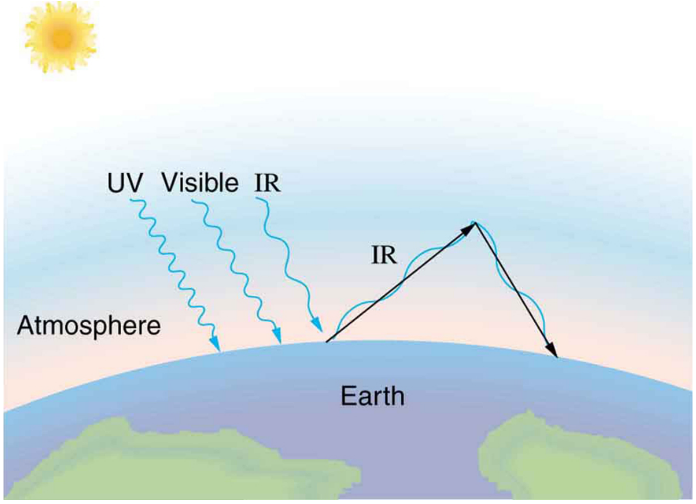
Equation:

$$= -99\text{ J/s} = -99\text{ W}.$$

Discussion

This value is a significant rate of heat transfer to the environment (note the minus sign), considering that a person at rest may produce energy at the rate of 125 W and that conduction and convection will also be transferring energy to the environment. Indeed, we would probably expect this person to feel cold. Clothing significantly reduces heat transfer to the environment by many methods, because clothing slows down both conduction and convection, and has a lower emissivity (especially if it is white) than skin.

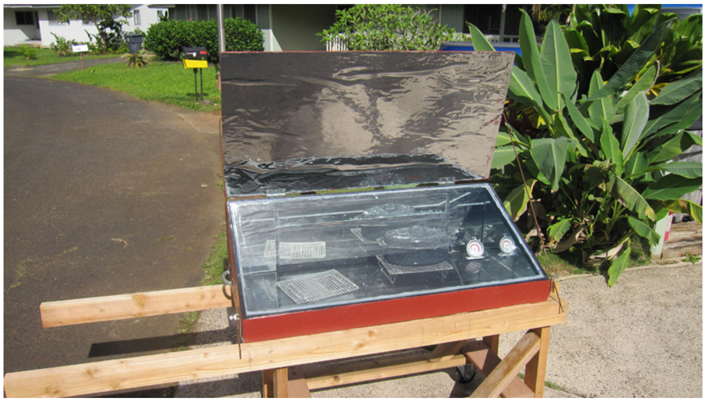
The Earth receives almost all its energy from radiation of the Sun and reflects some of it back into outer space. Because the Sun is hotter than the Earth, the net energy flux is from the Sun to the Earth. However, the rate of energy transfer is less than the equation for the radiative heat transfer would predict because the Sun does not fill the sky. The average emissivity (e) of the Earth is about 0.65, but the calculation of this value is complicated by the fact that the highly reflective cloud coverage varies greatly from day to day. There is a negative feedback (one in which a change produces an effect that opposes that change) between clouds and heat transfer; greater temperatures evaporate more water to form more clouds, which reflect more radiation back into space, reducing the temperature. The often mentioned **greenhouse effect** is directly related to the variation of the Earth's emissivity with radiation type (see the figure given below). The greenhouse effect is a natural phenomenon responsible for providing temperatures suitable for life on Earth. The Earth’s relatively constant temperature is a result of the energy balance between the incoming solar radiation and the energy radiated from the Earth. Most of the infrared radiation emitted from the Earth is absorbed by carbon dioxide (CO_2) and water (H_2O) in the atmosphere and then re-radiated back to the Earth or into outer space. Re-radiation back to the Earth maintains its surface temperature about $40^{\circ}C$ higher than it would be if there was no atmosphere, similar to the way glass increases temperatures in a greenhouse.



The greenhouse effect is a name given to the trapping of energy in the Earth’s atmosphere by a process similar to that used in greenhouses. The atmosphere, like window glass, is transparent to incoming visible radiation and most of the Sun’s infrared. These wavelengths are absorbed by the Earth and re-emitted as infrared. Since Earth’s temperature is much lower than that of the Sun, the infrared radiated by the Earth has a much longer wavelength. The atmosphere, like glass, traps these longer infrared rays, keeping the Earth warmer than it would otherwise be. The amount of trapping depends on concentrations of trace gases like carbon dioxide, and a change in the concentration of these gases is believed to affect the Earth’s surface temperature.

The greenhouse effect is also central to the discussion of global warming due to emission of carbon dioxide and methane (and other so-called greenhouse gases) into the Earth’s atmosphere from industrial production and farming. Changes in global climate could lead to more intense storms, precipitation changes (affecting agriculture), reduction in rain forest biodiversity, and rising sea levels.

Heating and cooling are often significant contributors to energy use in individual homes. Current research efforts into developing environmentally friendly homes quite often focus on reducing conventional heating and cooling through better building materials, strategically positioning windows to optimize radiation gain from the Sun, and opening spaces to allow convection. It is possible to build a zero-energy house that allows for comfortable living in most parts of the United States with hot and humid summers and cold winters.



This simple but effective solar cooker uses the greenhouse effect and reflective material to trap and retain solar energy. Made of inexpensive, durable materials, it saves money and labor, and is of particular economic value in energy-poor developing countries. (credit: E.B. Kauai)

Conversely, dark space is very cold, about $3\text{K}\ (-\ 454^{\text{o}}\text{F})$, so that the Earth radiates energy into the dark sky. Owing to the fact that clouds have lower emissivity than either oceans or land masses, they reflect some of the radiation back to the surface, greatly reducing heat transfer into dark space, just as they greatly reduce heat transfer into the atmosphere during the day. The rate of heat transfer from soil and grasses can be so rapid that frost may occur on clear summer evenings, even in warm latitudes.

Exercise:

Check Your Understanding

Problem:

What is the change in the rate of the radiated heat by a body at the temperature $T_1 = 20^{\text{o}}\text{C}$ compared to when the body is at the temperature $T_2 = 40^{\text{o}}\text{C}$?

Solution:

The radiated heat is proportional to the fourth power of the *absolute temperature*. Because $T_1 = 293\text{ K}$ and $T_2 = 313\text{ K}$, the rate of heat transfer increases by about 30 percent of the original rate.

Note:

Career Connection: Energy Conservation Consultation

The cost of energy is generally believed to remain very high for the foreseeable future. Thus, passive control of heat loss in both commercial and domestic housing will become increasingly important. Energy consultants measure and analyze the flow of energy into and out of houses and ensure that a healthy exchange of air is maintained inside the house. The job prospects for an energy consultant are strong.

Note:
Problem-Solving Strategies for the Methods of Heat Transfer

1. *Examine the situation to determine what type of heat transfer is involved.*
2. *Identify the type(s) of heat transfer—conduction, convection, or radiation.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is very useful.*
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns).*
5. *Solve the appropriate equation for the quantity to be determined (the unknown).*
6. For conduction, equation $\frac{Q}{t}=\frac{\mathrm{kA}(T_2-T_1)}{d}$ is appropriate. [link] lists thermal conductivities. For convection, determine the amount of matter moved and use equation $Q=\mathrm{mc}\Delta T$, to calculate the heat transfer involved in the temperature change of the fluid. If a phase change accompanies convection, equation $Q=\mathrm{mL_f}$ or $Q=\mathrm{mL_v}$ is appropriate to find the heat transfer involved in the phase change. [link] lists information relevant to phase change. For radiation, equation $\frac{Q_{\text{net}}}{t}=\sigma eA\left(T_2^4-T_1^4\right)$ gives the net heat transfer rate.
7. *Insert the knowns along with their units into the appropriate equation and obtain numerical solutions complete with units.*
8. *Check the answer to see if it is reasonable. Does it make sense?*

Summary

- Radiation is the rate of heat transfer through the emission or absorption of electromagnetic waves.

- The rate of heat transfer depends on the surface area and the fourth power of the absolute temperature:
  Equation:

  $$\frac{Q}{t} = \sigma e A T^4,$$

  where $\sigma = 5.67 \times 10^{-8}\ \mathrm{J/s \cdot m^2 \cdot K^4}$ is the Stefan-Boltzmann constant and e is the emissivity of the body. For a black body, $e = 1$ whereas a shiny white or perfect reflector has $e = 0$, with real objects having values of e between 1 and 0. The net rate of heat transfer by radiation is
  Equation:

  $$\frac{Q_{\text{net}}}{t} = \sigma e A \left(T_2^4 - T_1^4\right)$$

  where T_1 is the temperature of an object surrounded by an environment with uniform temperature T_2 and e is the emissivity of the *object*.

Conceptual Questions

Exercise:

Problem:

When watching a daytime circus in a large, dark-colored tent, you sense significant heat transfer from the tent. Explain why this occurs.

Exercise:

Problem:

Satellites designed to observe the radiation from cold (3 K) dark space have sensors that are shaded from the Sun, Earth, and Moon and that are cooled to very low temperatures. Why must the sensors be at low temperature?

Exercise:

Problem: Why are cloudy nights generally warmer than clear ones?

Exercise:

Problem:

Why are thermometers that are used in weather stations shielded from the sunshine? What does a thermometer measure if it is shielded from the sunshine and also if it is not?

Exercise:

Problem:

On average, would Earth be warmer or cooler without the atmosphere? Explain your answer.

Problems & Exercises

Exercise:

Problem:

At what net rate does heat radiate from a 275-m^2 black roof on a night when the roof's temperature is 30.0°C and the surrounding temperature is 15.0°C? The emissivity of the roof is 0.900.

Solution:

-21.7 kW
Note that the negative answer implies heat loss to the surroundings.

Exercise:

Problem:

(a) Cherry-red embers in a fireplace are at 850°C and have an exposed area of $0.200\ \text{m}^2$ and an emissivity of 0.980. The surrounding room has a temperature of 18.0°C. If 50% of the radiant energy enters the room, what is the net rate of radiant heat transfer in kilowatts? (b) Does your answer support the contention that most of the heat transfer into a room by a fireplace comes from infrared radiation?

Exercise:

Problem:

Radiation makes it impossible to stand close to a hot lava flow. Calculate the rate of heat transfer by radiation from $1.00\ \text{m}^2$ of 1200°C fresh lava into 30.0°C surroundings, assuming lava's emissivity is 1.00.

Solution:

$-266\ \text{kW}$

Exercise:

Problem:

(a) Calculate the rate of heat transfer by radiation from a car radiator at 110°C into a 50.0°C environment, if the radiator has an emissivity of 0.750 and a 1.20-m^2 surface area. (b) Is this a significant fraction of the heat transfer by an automobile engine? To answer this, assume a horsepower of $200\ \text{hp}\ (1.5\ \text{kW})$ and the efficiency of automobile engines as 25%.

Exercise:

Problem:

Find the net rate of heat transfer by radiation from a skier standing in the shade, given the following. She is completely clothed in white (head to foot, including a ski mask), the clothes have an emissivity of 0.200 and a surface temperature of 10.0°C, the surroundings are at -15.0°C, and her surface area is $1.60\ \text{m}^2$.

Solution:

$-36.0\ \text{W}$

Exercise:

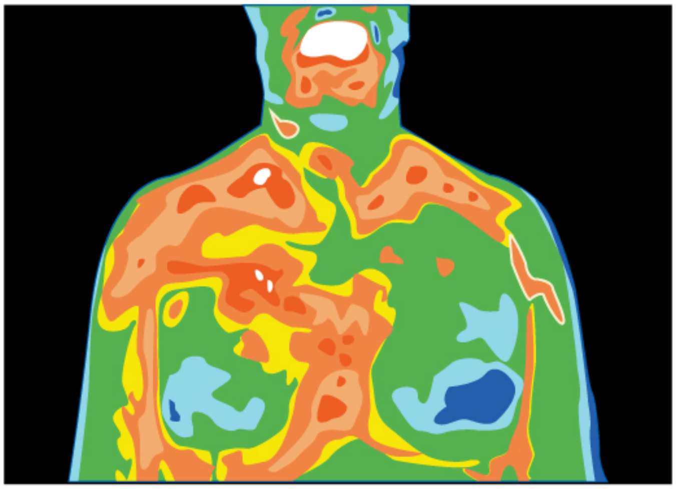
Problem:

Suppose you walk into a sauna that has an ambient temperature of 50.0°C. (a) Calculate the rate of heat transfer to you by radiation given your skin temperature is 37.0°C, the emissivity of skin is 0.98, and the surface area of your body is $1.50\ \text{m}^2$. (b) If all other forms of heat transfer are balanced (the net heat transfer is zero), at what rate will your body temperature increase if your mass is 75.0 kg?

Exercise:

Problem:

Thermography is a technique for measuring radiant heat and detecting variations in surface temperatures that may be medically, environmentally, or militarily meaningful.(a) What is the percent increase in the rate of heat transfer by radiation from a given area at a temperature of 34.0°C compared with that at 33.0°C, such as on a person’s skin? (b) What is the percent increase in the rate of heat transfer by radiation from a given area at a temperature of 34.0°C compared with that at 20.0°C, such as for warm and cool automobile hoods?



Artist’s rendition of a thermograph of a patient’s upper body, showing the distribution of heat represented by different colors.

Solution:

(a) 1.31%

(b) 20.5%

Exercise:

Problem:

The Sun radiates like a perfect black body with an emissivity of exactly 1. (a) Calculate the surface temperature of the Sun, given that it is a sphere with a 7.00×10^{8}-m radius that radiates 3.80×10^{26} W into 3-K space. (b) How much power does the Sun radiate per square meter of its surface? (c) How much power in watts per square meter is that value at the distance of Earth, 1.50×10^{11} m away? (This number is called the solar constant.)

Exercise:

Problem:

A large body of lava from a volcano has stopped flowing and is slowly cooling. The interior of the lava is at $1200^{\circ}\mathrm{C}$, its surface is at $450^{\circ}\mathrm{C}$, and the surroundings are at $27.0^{\circ}\mathrm{C}$. (a) Calculate the rate at which energy is transferred by radiation from $1.00\ \mathrm{m}^2$ of surface lava into the surroundings, assuming the emissivity is 1.00. (b) Suppose heat conduction to the surface occurs at the same rate. What is the thickness of the lava between the $450^{\circ}\mathrm{C}$ surface and the $1200^{\circ}\mathrm{C}$ interior, assuming that the lava's conductivity is the same as that of brick?

Solution:

(a) $-15.0\ \mathrm{kW}$

(b) 4.2 cm

Exercise:

Problem:

Calculate the temperature the entire sky would have to be in order to transfer energy by radiation at $1000\ \mathrm{W/m^2}$ —about the rate at which the Sun radiates when it is directly overhead on a clear day. This value is the effective temperature of the sky, a kind of average that takes account of the fact that the Sun occupies only a small part of the sky but is much hotter than the rest. Assume that the body receiving the energy has a temperature of $27.0^{\circ}\mathrm{C}$.

Exercise:

Problem:

(a) A shirtless rider under a circus tent feels the heat radiating from the sunlit portion of the tent. Calculate the temperature of the tent canvas based on the following information: The shirtless rider's skin temperature is 34.0^{o}C and has an emissivity of 0.970. The exposed area of skin is 0.400 m^2. He receives radiation at the rate of 20.0 W—half what you would calculate if the entire region behind him was hot. The rest of the surroundings are at 34.0^{o}C. (b) Discuss how this situation would change if the sunlit side of the tent was nearly pure white and if the rider was covered by a white tunic.

Solution:

(a) 48.5^{o}C

(b) A pure white object reflects more of the radiant energy that hits it, so a white tent would prevent more of the sunlight from heating up the inside of the tent, and the white tunic would prevent that heat which entered the tent from heating the rider. Therefore, with a white tent, the temperature would be lower than 48.5^{o}C, and the rate of radiant heat transferred to the rider would be less than 20.0 W.

Exercise:

Problem: **Integrated Concepts**

One 30.0^{o}C day the relative humidity is 75.0%, and that evening the temperature drops to 20.0^{o}C, well below the dew point. (a) How many grams of water condense from each cubic meter of air? (b) How much heat transfer occurs by this condensation? (c) What temperature increase could this cause in dry air?

Exercise:

Problem: **Integrated Concepts**

Large meteors sometimes strike the Earth, converting most of their kinetic energy into thermal energy. (a) What is the kinetic energy of a 10^9 kg meteor moving at 25.0 km/s? (b) If this meteor lands in a deep ocean and 80% of its kinetic energy goes into heating water, how many kilograms of water could it raise by $5.0^{\circ}\mathrm{C}$? (c) Discuss how the energy of the meteor is more likely to be deposited in the ocean and the likely effects of that energy.

Solution:

(a) $3 \times 10^{17}\ \mathrm{J}$

(b) $1 \times 10^{13}\ \mathrm{kg}$

(c) When a large meteor hits the ocean, it causes great tidal waves, dissipating large amount of its energy in the form of kinetic energy of the water.

Exercise:

Problem: Integrated Concepts

Frozen waste from airplane toilets has sometimes been accidentally ejected at high altitude. Ordinarily it breaks up and disperses over a large area, but sometimes it holds together and strikes the ground. Calculate the mass of $0^{\circ}\mathrm{C}$ ice that can be melted by the conversion of kinetic and gravitational potential energy when a $20.0\ \text{ kg}$ piece of frozen waste is released at 12.0 km altitude while moving at 250 m/s and strikes the ground at 100 m/s (since less than 20.0 kg melts, a significant mess results).

Exercise:

Problem: Integrated Concepts

(a) A large electrical power facility produces 1600 MW of "waste heat," which is dissipated to the environment in cooling towers by warming air flowing through the towers by $5.00^{\circ}\mathrm{C}$. What is the necessary flow rate of air in m^3/s? (b) Is your result consistent with the large cooling towers used by many large electrical power plants?

Solution:

(a) $3.44 \times 10^5\ \text{m}^3/\text{s}$

(b) This is equivalent to 12 million cubic feet of air per second. That is tremendous. This is too large to be dissipated by heating the air by only 5^oC. Many of these cooling towers use the circulation of cooler air over warmer water to increase the rate of evaporation. This would allow much smaller amounts of air necessary to remove such a large amount of heat because evaporation removes larger quantities of heat than was considered in part (a).

Exercise:

Problem: Integrated Concepts

(a) Suppose you start a workout on a Stairmaster, producing power at the same rate as climbing 116 stairs per minute. Assuming your mass is 76.0 kg and your efficiency is 20.0%, how long will it take for your body temperature to rise 1.00^oC if all other forms of heat transfer in and out of your body are balanced? (b) Is this consistent with your experience in getting warm while exercising?

Exercise:

Problem: Integrated Concepts

A 76.0-kg person suffering from hypothermia comes indoors and shivers vigorously. How long does it take the heat transfer to increase the person’s body temperature by 2.00^oC if all other forms of heat transfer are balanced?

Solution:

20.9 min

Exercise:

Problem: Integrated Concepts

In certain large geographic regions, the underlying rock is hot. Wells can be drilled and water circulated through the rock for heat transfer for the generation of electricity. (a) Calculate the heat transfer that can be extracted by cooling $1.00\ \text{km}^3$ of granite by 100^{o}C. (b) How long will this take if heat is transferred at a rate of 300 MW, assuming no heat transfers back into the 1.00 km of rock by its surroundings?

Exercise:

Problem: Integrated Concepts

Heat transfers from your lungs and breathing passages by evaporating water. (a) Calculate the maximum number of grams of water that can be evaporated when you inhale 1.50 L of 37^{o}C air with an original relative humidity of 40.0%. (Assume that body temperature is also 37^{o}C.) (b) How many joules of energy are required to evaporate this amount? (c) What is the rate of heat transfer in watts from this method, if you breathe at a normal resting rate of 10.0 breaths per minute?

Solution:

(a) 3.96×10^{-2} g

(b) 96.2 J

(c) 16.0 W

Exercise:

Problem: Integrated Concepts

(a) What is the temperature increase of water falling 55.0 m over Niagara Falls? (b) What fraction must evaporate to keep the temperature constant?

Exercise:

Problem: **Integrated Concepts**

Hot air rises because it has expanded. It then displaces a greater volume of cold air, which increases the buoyant force on it. (a) Calculate the ratio of the buoyant force to the weight of 50.0^{o}C air surrounded by 20.0^{o}C air. (b) What energy is needed to cause 1.00 m^3 of air to go from 20.0^{o}C to 50.0^{o}C? (c) What gravitational potential energy is gained by this volume of air if it rises 1.00 m? Will this cause a significant cooling of the air?

Solution:

(a) 1.102

(b) $2.79\times 10^4\text{ J}$

(c) 12.6 J. This will not cause a significant cooling of the air because it is much less than the energy found in part (b), which is the energy required to warm the air from 20.0^{o}C to 50.0^{o}C.

Exercise:

Problem: **Unreasonable Results**

(a) What is the temperature increase of an 80.0 kg person who consumes 2500 kcal of food in one day with 95.0% of the energy transferred as heat to the body? (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Solution:

(a) 36^{o}C

(b) Any temperature increase greater than about 3^{o}C would be unreasonably large. In this case the final temperature of the person would rise to $73^{\text{o}}\text{C}\ (163^{\text{o}}\text{F})$.

(c) The assumption of 95% heat retention is unreasonable.

Exercise:

Problem: Unreasonable Results

A slightly deranged Arctic inventor surrounded by ice thinks it would be much less mechanically complex to cool a car engine by melting ice on it than by having a water-cooled system with a radiator, water pump, antifreeze, and so on. (a) If 80.0% of the energy in 1.00 gal of gasoline is converted into "waste heat" in a car engine, how many kilograms of $0^{\circ}\mathrm{C}$ ice could it melt? (b) Is this a reasonable amount of ice to carry around to cool the engine for 1.00 gal of gasoline consumption? (c) What premises or assumptions are unreasonable?

Exercise:

Problem: Unreasonable Results

(a) Calculate the rate of heat transfer by conduction through a window with an area of $1.00\ \mathrm{m}^2$ that is 0.750 cm thick, if its inner surface is at $22.0^{\circ}\mathrm{C}$ and its outer surface is at $35.0^{\circ}\mathrm{C}$. (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Solution:

(a) 1.46 kW

(b) Very high power loss through a window. An electric heater of this power can keep an entire room warm.

(c) The surface temperatures of the window do not differ by as great an amount as assumed. The inner surface will be warmer, and the outer surface will be cooler.

Exercise:

Problem: Unreasonable Results

A meteorite 1.20 cm in diameter is so hot immediately after penetrating the atmosphere that it radiates 20.0 kW of power. (a) What is its temperature, if the surroundings are at 20.0°C and it has an emissivity of 0.800? (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Exercise:

Problem: Construct Your Own Problem

Consider a new model of commercial airplane having its brakes tested as a part of the initial flight permission procedure. The airplane is brought to takeoff speed and then stopped with the brakes alone. Construct a problem in which you calculate the temperature increase of the brakes during this process. You may assume most of the kinetic energy of the airplane is converted to thermal energy in the brakes and surrounding materials, and that little escapes. Note that the brakes are expected to become so hot in this procedure that they ignite and, in order to pass the test, the airplane must be able to withstand the fire for some time without a general conflagration.

Exercise:

Problem: Construct Your Own Problem

Consider a person outdoors on a cold night. Construct a problem in which you calculate the rate of heat transfer from the person by all three heat transfer methods. Make the initial circumstances such that at rest the person will have a net heat transfer and then decide how much physical activity of a chosen type is necessary to balance the rate of heat transfer. Among the things to consider are the size of the person, type of clothing, initial metabolic rate, sky conditions, amount of water evaporated, and volume of air breathed. Of course, there are many other factors to consider and your instructor may wish to guide you in the assumptions made as well as the detail of analysis and method of presenting your results.

Glossary

emissivity
    measure of how well an object radiates

greenhouse effect
    warming of the Earth that is due to gases such as carbon dioxide and methane that absorb infrared radiation from the Earth’s surface and reradiate it in all directions, thus sending a fraction of it back toward the surface of the Earth

net rate of heat transfer by radiation
    is $\frac{Q_{\text{net}}}{t} = \sigma e A \left(T_2^4 - T_1^4 \right)$

radiation
    energy transferred by electromagnetic waves directly as a result of a temperature difference

Stefan-Boltzmann law of radiation
    $\frac{Q}{t} = \sigma e A T^4$, where σ is the Stefan-Boltzmann constant, A is the surface area of the object, T is the absolute temperature, and e is the emissivity

Introduction to Thermodynamics class="introduction"

A steam engine uses heat transfer to do work. Tourists regularly ride this narrow-gauge steam engine train near the San Juan Skyway in Durango, Colorado, part of the National Scenic Byways Program. (credit: Dennis Adams)

484
KEEP OFF
KEEP OFF

Heat transfer is energy in transit, and it can be used to do work. It can also be converted to any other form of energy. A car engine, for example, burns fuel for heat transfer into a gas. Work is done by the gas as it exerts a force through a distance, converting its energy into a variety of other forms—into the car's kinetic or gravitational potential energy; into electrical energy to run the spark plugs, radio, and lights; and back into stored energy in the car's battery. But most of the heat transfer produced from burning fuel in the engine does not do work on the gas. Rather, the energy is released into the environment, implying that the engine is quite inefficient.

It is often said that modern gasoline engines cannot be made to be significantly more efficient. We hear the same about heat transfer to electrical energy in large power stations, whether they are coal, oil, natural gas, or nuclear powered. Why is that the case? Is the inefficiency caused by design problems that could be solved with better engineering and superior materials? Is it part of some money-making conspiracy by those who sell energy? Actually, the truth is more interesting, and reveals much about the nature of heat transfer.

Basic physical laws govern how heat transfer for doing work takes place and place insurmountable limits onto its efficiency. This chapter will explore these laws as well as many applications and concepts associated with them. These topics are part of *thermodynamics*—the study of heat transfer and its relationship to doing work.

The First Law of Thermodynamics

- Define the first law of thermodynamics.
- Describe how conservation of energy relates to the first law of thermodynamics.
- Identify instances of the first law of thermodynamics working in everyday situations, including biological metabolism.
- Calculate changes in the internal energy of a system, after accounting for heat transfer and work done.

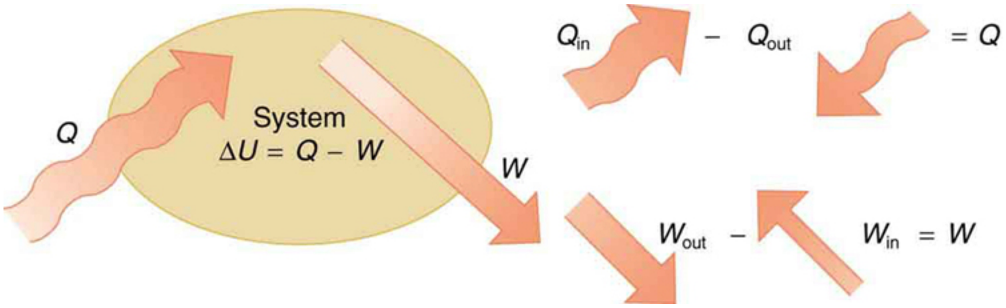
This boiling tea kettle represents energy in motion. The water in the kettle is turning to water vapor because heat is being transferred from the stove to the kettle. As the entire system gets hotter, work is done—from the evaporation of the water to the whistling of the kettle. (credit: Gina Hamilton)

If we are interested in how heat transfer is converted into doing work, then the conservation of energy principle is important. The first law of thermodynamics applies the conservation of energy principle to systems where heat transfer and doing work are the methods of transferring energy into and out of the system. The **first law of thermodynamics** states that the change in internal energy of a system equals the net heat transfer *into* the system minus the net work done *by* the system. In equation form, the first law of thermodynamics is

Equation:

$$\Delta U = Q - W.$$

Here ΔU is the *change in internal energy* U of the system. Q is the *net heat transferred into the system*—that is, Q is the sum of all heat transfer into and out of the system. W is the *net work done by the system*—that is, W is the sum of all work done on or by the system. We use the following sign conventions: if Q is positive, then there is a net heat transfer into the system; if W is positive, then there is net work done by the system. So positive Q adds energy to the system and positive W takes energy from the system. Thus $\Delta U = Q - W$. Note also that if more heat transfer into the system occurs than work done, the difference is stored as internal energy. Heat engines are a good example of this—heat transfer into them takes place so that they can do work. (See [link].) We will now examine Q, W, and ΔU further.



The first law of thermodynamics is the conservation-of-energy principle stated for a system where heat and work are the methods of transferring energy for a system in thermal equilibrium. Q represents the net heat transfer—it is the sum of all heat transfers into and out of the system. Q is positive for net heat transfer *into* the system. W is the total work done on and by the system. W is positive when more work is done *by* the system than on it. The change in the internal energy of the system, ΔU, is related to heat and work by the first law of thermodynamics, $\Delta U = Q - W$.

Note:
Making Connections: Law of Thermodynamics and Law of Conservation of Energy
The first law of thermodynamics is actually the law of conservation of energy stated in a form most useful in thermodynamics. The first law gives the relationship between heat transfer, work done, and the change in internal energy of a system.

Heat *Q* and Work *W*

Heat transfer (Q) and doing work (W) are the two everyday means of bringing energy into or taking energy out of a system. The processes are quite different. Heat transfer, a less organized process, is driven by temperature differences. Work, a quite organized process, involves a macroscopic force exerted through a distance. Nevertheless, heat and work can produce identical results.For example, both can cause a temperature increase. Heat transfer into a system, such as when the Sun warms the air in a bicycle tire, can increase its temperature, and so can work done on the system, as when the bicyclist pumps air into the tire. Once the temperature increase has occurred, it is impossible to tell whether it was caused by heat transfer or by doing work. This uncertainty is an important point. Heat transfer and work are both energy in transit—neither is stored as such in a system. However, both can change the internal energy U of a system. Internal energy is a form of energy completely different from either heat or work.

Internal Energy U

We can think about the internal energy of a system in two different but consistent ways. The first is the atomic and molecular view, which examines the system on the atomic and molecular scale. The **internal energy** U of a system is the sum of the kinetic and potential energies of its atoms and molecules. Recall that kinetic plus potential energy is called mechanical energy. Thus internal energy is the sum of atomic and molecular mechanical energy. Because it is impossible to keep track of all individual atoms and molecules, we must deal with averages and distributions. A second way to view the internal energy of a system is in terms of its macroscopic characteristics, which are very similar to atomic and molecular average values.

Macroscopically, we define the change in internal energy ΔU to be that given by the first law of thermodynamics:

Equation:

$$\Delta U = Q - W.$$

Many detailed experiments have verified that $\Delta U = Q - W$, where ΔU is the change in total kinetic and potential energy of all atoms and molecules in a system. It has also been determined experimentally that the internal energy U of a system depends only on the state of the system and *not how it reached that state*. More specifically, U is found to be a function of a few macroscopic quantities (pressure, volume, and temperature, for example), independent of past history such as whether there has been heat transfer or work done. This independence means that if we know the state of a system, we can calculate changes in its internal energy U from a few macroscopic variables.

Note:

Making Connections: Macroscopic and Microscopic

In thermodynamics, we often use the macroscopic picture when making calculations of how a system behaves, while the atomic and molecular picture gives underlying explanations in terms of averages and distributions. We shall see this again in later sections of this chapter. For example, in the topic of entropy, calculations will be made using the atomic and molecular view.

To get a better idea of how to think about the internal energy of a system, let us examine a system going from State 1 to State 2. The system has internal energy U_1 in State 1, and it has internal energy U_2 in State 2, no matter how it got to either state. So the change in internal energy $\Delta U = U_2 - U_1$ is independent of what caused the change. In other words, *ΔU is independent of path*. By path, we mean the method of getting from the starting point to the ending point. Why is this independence important? Note that $\Delta U = Q - W$. Both Q and W *depend on path*, but ΔU does not. This path independence means that internal energy U is easier to consider than either heat transfer or work done.

Example:

Calculating Change in Internal Energy: The Same Change in U is Produced by Two Different Processes

(a) Suppose there is heat transfer of 40.00 J to a system, while the system does 10.00 J of work. Later, there is heat transfer of 25.00 J out of the system while 4.00 J of work is done on the system. What is the net change in internal energy of the system?

(b) What is the change in internal energy of a system when a total of 150.00 J of heat transfer occurs out of (from) the system and 159.00 J of work is done on the system? (See [link]).

Strategy

In part (a), we must first find the net heat transfer and net work done from the given information. Then the first law of thermodynamics

$(\Delta U = Q - W)$ can be used to find the change in internal energy. In part (b), the net heat transfer and work done are given, so the equation can be used directly.

Solution for (a)

The net heat transfer is the heat transfer into the system minus the heat transfer out of the system, or

Equation:

$$Q = 40.00\text{ J} - 25.00\text{ J} = 15.00\text{ J}.$$

Similarly, the total work is the work done by the system minus the work done on the system, or

Equation:

$$W = 10.00\text{ J} - 4.00\text{ J} = 6.00\text{ J}.$$

Thus the change in internal energy is given by the first law of thermodynamics:

Equation:

$$\Delta U = Q - W = 15.00\text{ J} - 6.00\text{ J} = 9.00\text{ J}.$$

We can also find the change in internal energy for each of the two steps. First, consider 40.00 J of heat transfer in and 10.00 J of work out, or

Equation:

$$\Delta U_1 = Q_1 - W_1 = 40.00\text{ J} - 10.00\text{ J} = 30.00\text{ J}.$$

Now consider 25.00 J of heat transfer out and 4.00 J of work in, or

Equation:

$$\Delta U_2 = Q_2 - W_2 = -25.00\text{ J} - (-4.00\text{ J}) = -21.00\text{ J}.$$

The total change is the sum of these two steps, or

Equation:

$$\Delta U = \Delta U_1 + \Delta U_2 = 30.00\text{ J} + (-21.00\text{ J}) = 9.00\text{ J}.$$

Discussion on (a)

No matter whether you look at the overall process or break it into steps, the change in internal energy is the same.

Solution for (b)

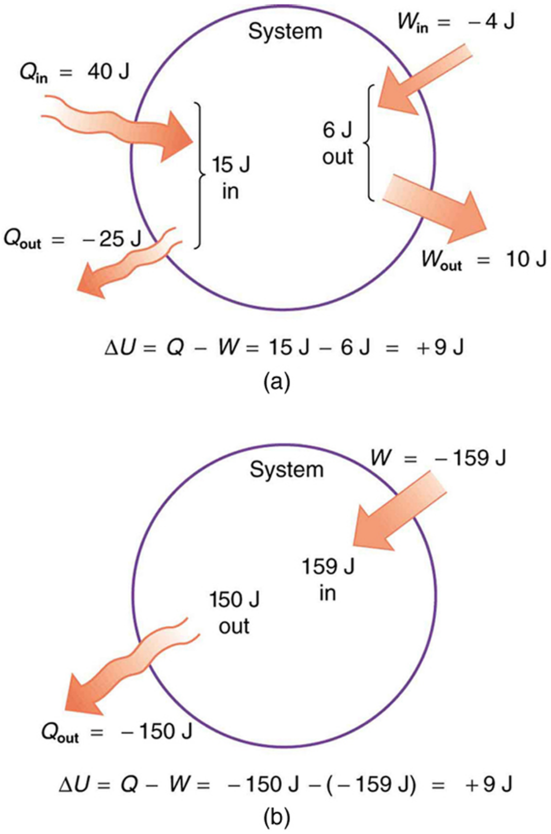
Here the net heat transfer and total work are given directly to be $Q = -150.00\text{ J}$ and $W = -159.00\text{ J}$, so that

Equation:

$$\Delta U = Q - W = -150.00\text{ J} - (-159.00\text{ J}) = 9.00\text{ J}.$$

Discussion on (b)

A very different process in part (b) produces the same 9.00-J change in internal energy as in part (a). Note that the change in the system in both parts is related to ΔU and not to the individual Qs or Ws involved. The system ends up in the *same* state in both (a) and (b). Parts (a) and (b) present two different paths for the system to follow between the same starting and ending points, and the change in internal energy for each is the same—it is independent of path.


System
W_{in} = −4 J
Q_{in} = 40 J
6 J out
15 J in
Q_{out} = −25 J
W_{out} = 10 J
$\Delta U = Q - W = 15\text{ J} - 6\text{ J} = +9\text{ J}$
(a)
System
W = −159 J
159 J in
150 J out
Q_{out} = −150 J
$\Delta U = Q - W = -150\text{ J} - (-159\text{ J}) = +9\text{ J}$
(b)

Two different processes produce the same change in a system. (a) A total of 15.00 J of heat transfer occurs into the system, while work takes out a total of 6.00 J. The change in internal energy is $\Delta U = Q - W = 9.00\text{ J}$. (b) Heat transfer removes 150.00 J from the system while work puts 159.00 J into it, producing an increase of 9.00 J in internal energy. If the system starts out in the same state in (a) and (b), it will end up in the same final state in either case—its final state is related to internal

energy, not how that energy
was acquired.

Human Metabolism and the First Law of Thermodynamics

Human metabolism is the conversion of food into heat transfer, work, and stored fat. Metabolism is an interesting example of the first law of thermodynamics in action. We now take another look at these topics via the first law of thermodynamics. Considering the body as the system of interest, we can use the first law to examine heat transfer, doing work, and internal energy in activities ranging from sleep to heavy exercise. What are some of the major characteristics of heat transfer, doing work, and energy in the body? For one, body temperature is normally kept constant by heat transfer to the surroundings. This means Q is negative. Another fact is that the body usually does work on the outside world. This means W is positive. In such situations, then, the body loses internal energy, since $\Delta U = Q - W$ is negative.

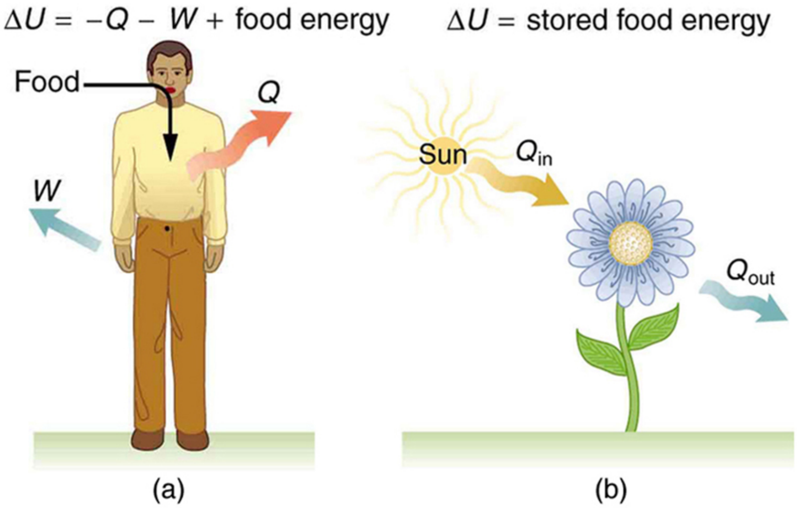
Now consider the effects of eating. Eating increases the internal energy of the body by adding chemical potential energy (this is an unromantic view of a good steak). The body *metabolizes* all the food we consume. Basically, metabolism is an oxidation process in which the chemical potential energy of food is released. This implies that food input is in the form of work. Food energy is reported in a special unit, known as the Calorie. This energy is measured by burning food in a calorimeter, which is how the units are determined.

In chemistry and biochemistry, one calorie (spelled with a *lowercase* c) is defined as the energy (or heat transfer) required to raise the temperature of one gram of pure water by one degree Celsius. Nutritionists and weight-watchers tend to use the *dietary* calorie, which is frequently called a Calorie (spelled with a *capital* C). One food Calorie is the energy needed to raise the temperature of one *kilogram* of water by one degree Celsius. This means that one dietary Calorie is equal to one kilocalorie for the chemist, and one must be careful to avoid confusion between the two.

Again, consider the internal energy the body has lost. There are three places this internal energy can go—to heat transfer, to doing work, and to stored fat (a tiny fraction also goes to cell repair and growth). Heat transfer and doing work take internal energy out of the body, and food puts it back. If you eat just the right amount of food, then your average internal energy remains constant. Whatever you lose to heat transfer and doing work is replaced by food, so that, in the long run, $\Delta U = 0$. If you overeat repeatedly, then ΔU is always positive, and your body stores this extra internal energy as fat. The reverse is true if you eat too little. If ΔU is negative for a few days, then the body metabolizes its own fat to maintain body temperature and do work that takes energy from the body. This process is how dieting produces weight loss.

Life is not always this simple, as any dieter knows. The body stores fat or metabolizes it only if energy intake changes for a period of several days. Once you have been on a major diet, the next one is less successful because your body alters the way it responds to low energy intake. Your basal metabolic rate (BMR) is the rate at which food is converted into heat transfer and work done while the body is at complete rest. The body adjusts its basal metabolic rate to partially compensate for over-eating or under-eating. The body will decrease the metabolic rate rather than eliminate its own fat to replace lost food intake. You will chill more easily and feel less energetic as a result of the lower metabolic rate, and you will not lose weight as fast as before. Exercise helps to lose weight, because it produces both heat transfer from your body and work, and raises your metabolic rate even when you are at rest. Weight loss is also aided by the quite low efficiency of the body in converting internal energy to work, so that the loss of internal energy resulting from doing work is much greater than the work done.It should be noted, however, that living systems are not in thermalequilibrium.

The body provides us with an excellent indication that many thermodynamic processes are *irreversible*. An irreversible process can go in one direction but not the reverse, under a given set of conditions. For example, although body fat can be converted to do work and produce heat transfer, work done on the body and heat transfer into it cannot be converted to body fat. Otherwise, we could skip lunch by sunning ourselves or by walking down stairs. Another example of an irreversible thermodynamic process is photosynthesis. This process is the intake of one form of energy—light—by plants and its conversion to chemical potential energy. Both applications of the first law of thermodynamics are illustrated in [link]. One great advantage of conservation laws such as the first law of thermodynamics is that they accurately describe the beginning and ending points of complex processes, such as metabolism and photosynthesis, without regard to the complications in between. [link] presents a summary of terms relevant to the first law of thermodynamics.


$\Delta U = -Q - W + \text{food energy}$
$\Delta U = \text{stored food energy}$
Food
Q
W
Sun
Q_{in}
Q_{out}
(a)
(b)

(a) The first law of thermodynamics applied to metabolism. Heat transferred out of the body (Q) and work done by the body (W) remove internal energy, while food intake replaces it. (Food intake may be considered as work done on the body.) (b) Plants convert part of the radiant heat transfer in sunlight to stored chemical energy, a process called photosynthesis.

Term	Definition
U	Internal energy—the sum of the kinetic and potential energies of a system's atoms and molecules. Can be divided into many subcategories, such as thermal and chemical energy. Depends only on the state of a system (such as its P, V, and T), not on how the energy entered the system. Change in internal energy is path independent.
Q	Heat—energy transferred because of a temperature difference. Characterized by random molecular motion. Highly dependent on path. Q entering a system is positive.
W	Work—energy transferred by a force moving through a distance. An organized, orderly process. Path dependent. W done by a system (either against an external force or to increase the volume of the system) is positive.

Summary of Terms for the First Law of Thermodynamics, $\Delta U=Q-W$

Section Summary

- The first law of thermodynamics is given as $\Delta U = Q - W$, where ΔU is the change in internal energy of a system, Q is the net heat transfer (the sum of all heat transfer into and out of the system), and W is the net work done (the sum of all work done on or by the system).
- Both Q and W are energy in transit; only ΔU represents an independent quantity capable of being stored.

- The internal energy U of a system depends only on the state of the system and not how it reached that state.
- Metabolism of living organisms, and photosynthesis of plants, are specialized types of heat transfer, doing work, and internal energy of systems.

Conceptual Questions

Exercise:

Problem:

Describe the photo of the tea kettle at the beginning of this section in terms of heat transfer, work done, and internal energy. How is heat being transferred? What is the work done and what is doing it? How does the kettle maintain its internal energy?

Exercise:

Problem:

The first law of thermodynamics and the conservation of energy, as discussed in Conservation of Energy, are clearly related. How do they differ in the types of energy considered?

Exercise:

Problem:

Heat transfer Q and work done W are always energy in transit, whereas internal energy U is energy stored in a system. Give an example of each type of energy, and state specifically how it is either in transit or resides in a system.

Exercise:

Problem:

How do heat transfer and internal energy differ? In particular, which can be stored as such in a system and which cannot?

Exercise:

Problem:

If you run down some stairs and stop, what happens to your kinetic energy and your initial gravitational potential energy?

Exercise:

Problem:

Give an explanation of how food energy (calories) can be viewed as molecular potential energy (consistent with the atomic and molecular definition of internal energy).

Exercise:

Problem:

Identify the type of energy transferred to your body in each of the following as either internal energy, heat transfer, or doing work: (a) basking in sunlight; (b) eating food; (c) riding an elevator to a higher floor.

Problems & Exercises

Exercise:

Problem:

What is the change in internal energy of a car if you put 12.0 gal of gasoline into its tank? The energy content of gasoline is $1.3 \times 10^8 \text{ J/gal}$. All other factors, such as the car's temperature, are constant.

Solution:

$1.6 \times 10^9 \text{ J}$

Exercise:

Problem:

How much heat transfer occurs from a system, if its internal energy decreased by 150 J while it was doing 30.0 J of work?

Exercise:

Problem:

A system does $1.80\times 10^8\ \text{J}$ of work while $7.50\times 10^8\ \text{J}$ of heat transfer occurs to the environment. What is the change in internal energy of the system assuming no other changes (such as in temperature or by the addition of fuel)?

Solution:

$-9.30\times 10^8\ \text{J}$

Exercise:

Problem:

What is the change in internal energy of a system which does $4.50\times 10^5\ \text{J}$ of work while $3.00\times 10^6\ \text{J}$ of heat transfer occurs into the system, and $8.00\times 10^6\ \text{J}$ of heat transfer occurs to the environment?

Exercise:

Problem:

Suppose a woman does 500 J of work and 9500 J of heat transfer occurs into the environment in the process. (a) What is the decrease in her internal energy, assuming no change in temperature or consumption of food? (That is, there is no other energy transfer.) (b) What is her efficiency?

Solution:

(a) $-1.0\times 10^4\ \text{J}$, or $-2.39\ \text{kcal}$

(b) 5.00%

Exercise:

Problem:

(a) How much food energy will a man metabolize in the process of doing 35.0 kJ of work with an efficiency of 5.00%? (b) How much heat transfer occurs to the environment to keep his temperature constant? Explicitly show how you follow the steps in the Problem-Solving Strategy for thermodynamics found in Problem-Solving Strategies for Thermodynamics.

Exercise:

Problem:

(a) What is the average metabolic rate in watts of a man who metabolizes 10,500 kJ of food energy in one day? (b) What is the maximum amount of work in joules he can do without breaking down fat, assuming a maximum efficiency of 20.0%? (c) Compare his work output with the daily output of a 187-W (0.250-horsepower) motor.

Solution:

(a) 122 W

(b) 2.10×10^6 J

(c) Work done by the motor is 1.61×10^7 J ;thus the motor produces 7.67 times the work done by the man

Exercise:

Problem:

(a) How long will the energy in a 1470-kJ (350-kcal) cup of yogurt last in a woman doing work at the rate of 150 W with an efficiency of 20.0% (such as in leisurely climbing stairs)? (b) Does the time found in part (a) imply that it is easy to consume more food energy than you can reasonably expect to work off with exercise?

Exercise:

Problem:

(a) A woman climbing the Washington Monument metabolizes 6.00×10^2 kJ of food energy. If her efficiency is 18.0%, how much heat transfer occurs to the environment to keep her temperature constant? (b) Discuss the amount of heat transfer found in (a). Is it consistent with the fact that you quickly warm up when exercising?

Solution:

(a) 492 kJ

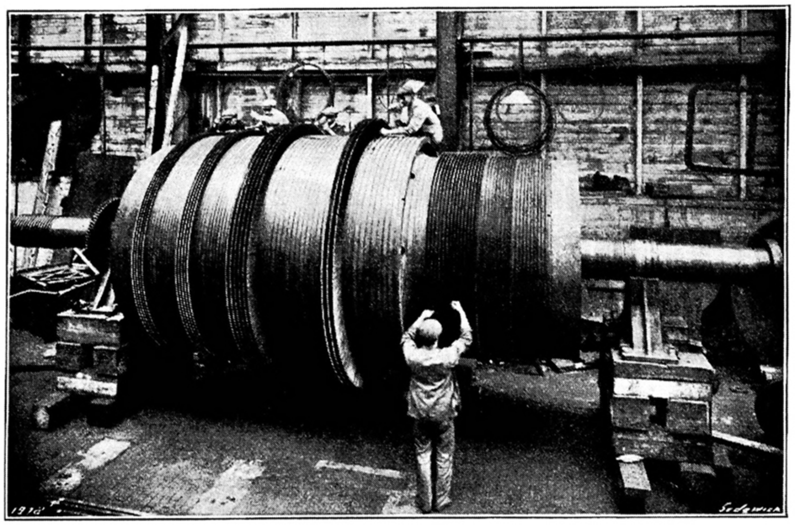
(b) This amount of heat is consistent with the fact that you warm quickly when exercising. Since the body is inefficient, the excess heat produced must be dissipated through sweating, breathing, etc.

Glossary

first law of thermodynamics	states that the change in internal energy of a system equals the net heat transfer *into* the system minus the net work done *by* the system
internal energy	the sum of the kinetic and potential energies of a system’s atoms and molecules
human metabolism	conversion of food into heat transfer, work, and stored fat

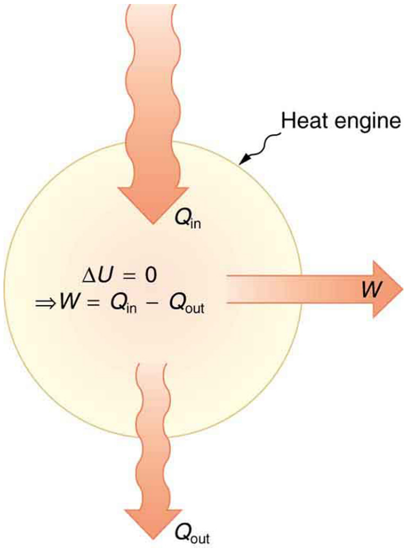
The First Law of Thermodynamics and Some Simple Processes

- Describe the processes of a simple heat engine.
- Explain the differences among the simple thermodynamic processes—isobaric, isochoric, isothermal, and adiabatic.
- Calculate total work done in a cyclical thermodynamic process.

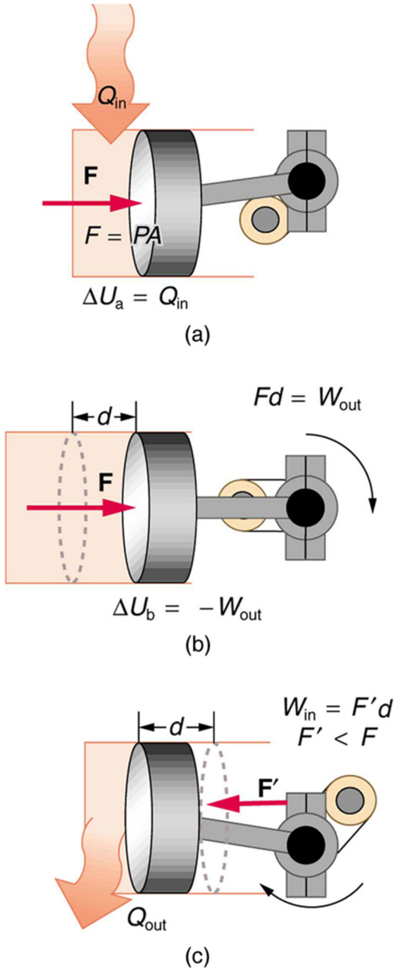


Beginning with the Industrial Revolution, humans have harnessed power through the use of the first law of thermodynamics, before we even understood it completely. This photo, of a steam engine at the Turbinia Works, dates from 1911, a mere 61 years after the first explicit statement of the first law of thermodynamics by Rudolph Clausius. (credit: public domain; author unknown)

One of the most important things we can do with heat transfer is to use it to do work for us. Such a device is called a **heat engine**. Car engines and steam turbines that generate electricity are examples of heat engines. [link] shows schematically how the first law of thermodynamics applies to the typical heat engine.


Heat engine
Q_{in}
$\Delta U = 0$
$\Rightarrow W = Q_{in} - Q_{out}$
W
Q_{out}

Schematic representation of a heat engine, governed, of course, by the first law of thermodynamics. It is impossible to devise a system where $Q_{\text{out}} = 0$, that is, in which no heat transfer occurs to the environment.


Q_{in}
F
$F = PA$
$\Delta U_a = Q_{in}$
(a)
$Fd = W_{out}$
d
F
$\Delta U_b = -W_{out}$
(b)
$W_{in} = F'd$
$F' < F$
d
F′
Q_{out}
(c)

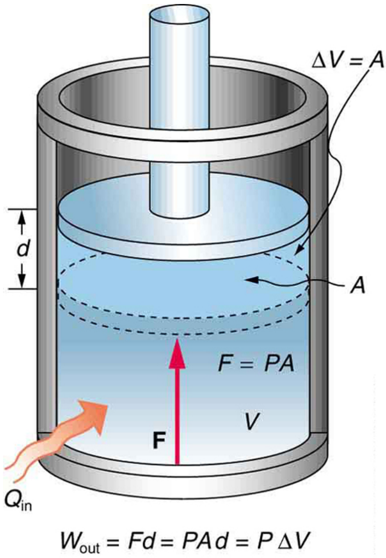
(a) Heat transfer to the gas in a cylinder increases the internal energy of the gas, creating higher pressure and temperature. (b) The force exerted on the movable cylinder does work as the gas expands. Gas pressure and temperature decrease when it expands, indicating that the gas's internal energy has been decreased by doing work. (c) Heat transfer to the environment further reduces pressure in the gas so that the piston can be more easily returned to its starting position.

The illustrations above show one of the ways in which heat transfer does work. Fuel combustion produces heat transfer to a gas in a cylinder, increasing the pressure of the gas and thereby the force it exerts on a movable piston. The gas does work on the outside world, as this force moves the piston through some distance. Heat transfer to the gas cylinder results in work being done. To repeat this process, the piston needs to be returned to its starting point. Heat transfer now occurs from the gas to the surroundings so that its pressure decreases, and a force is exerted by the surroundings to push the piston back through some distance. Variations of this process are employed daily in hundreds of millions of heat engines. We will examine heat engines in detail in the next section. In this section, we consider some of the simpler underlying processes on which heat engines are based.

PV Diagrams and their Relationship to Work Done on or by a Gas

A process by which a gas does work on a piston at constant pressure is called an **isobaric process**. Since the pressure is constant, the force exerted is constant and the work done is given as
Equation:

$$P\Delta V.$$



ΔV = Ad
d
A
F = PA
V
F
Q_{in}
$W_{out} = Fd = PAd = P\Delta V$

An isobaric expansion of a gas requires heat transfer to keep the pressure constant. Since pressure is constant, the work done is $P\Delta V$.

Equation:

$$W=\text{Fd}$$

See the symbols as shown in [link]. Now $F=\text{PA}$, and so

Equation:

$$W=\text{PAd}.$$

Because the volume of a cylinder is its cross-sectional area A times its length d, we see that $\mathrm{Ad} = \Delta V$, the change in volume; thus,

Equation:

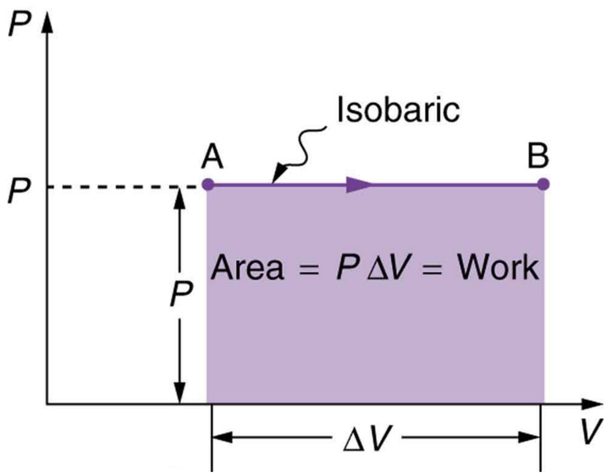
$$W = P\Delta V \text{ (isobaric process)}.$$

Note that if ΔV is positive, then W is positive, meaning that work is done *by* the gas on the outside world.

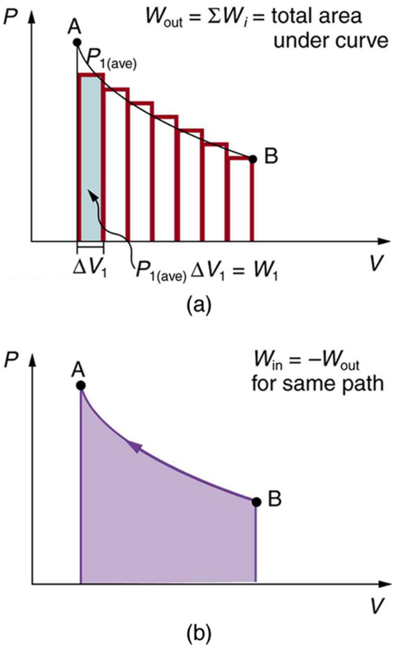
(Note that the pressure involved in this work that we’ve called P is the pressure of the gas *inside* the tank. If we call the pressure outside the tank P_{ext}, an expanding gas would be working *against* the external pressure; the work done would therefore be $W = -P_{\text{ext}}\Delta V$ (isobaric process). Many texts use this definition of work, and not the definition based on internal pressure, as the basis of the First Law of Thermodynamics. This definition reverses the sign conventions for work, and results in a statement of the first law that becomes $\Delta U = Q + W$.)

It is not surprising that $W = P\Delta V$, since we have already noted in our treatment of fluids that pressure is a type of potential energy per unit volume and that pressure in fact has units of energy divided by volume. We also noted in our discussion of the ideal gas law that PV has units of energy. In this case, some of the energy associated with pressure becomes work.

[link] shows a graph of pressure versus volume (that is, a PV diagram for an isobaric process. You can see in the figure that the work done is the area under the graph. This property of PV diagrams is very useful and broadly applicable: *the work done on or by a system in going from one state to another equals the area under the curve on a* PV *diagram*.


P
Isobaric
A
B
P
Area = P ΔV = Work
P
ΔV
V

A graph of pressure versus volume for a constant-pressure, or isobaric, process, such as the one shown in [link]. The area under the curve equals the work done by the gas, since $W = P\Delta V.$


P
A
$W_{out} = \Sigma W_i$ = total area under curve
$P_{1(ave)}$
B
ΔV_1
$P_{1(ave)} \Delta V_1 = W_1$
V
P
A
$W_{in} = -W_{out}$ for same path
B
V

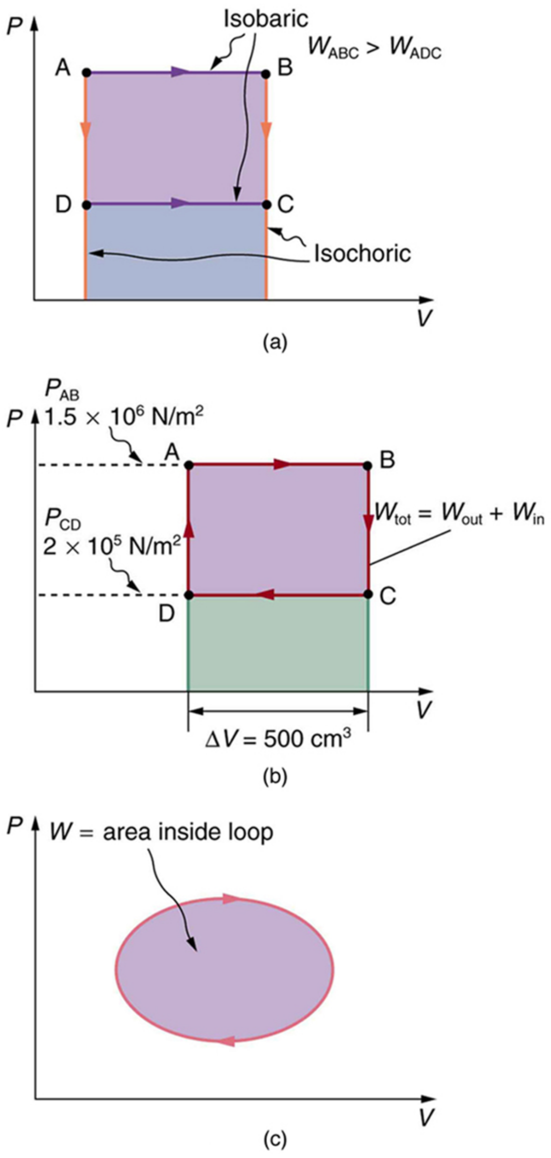
(a)

(b)

(a) A PV diagram in which pressure varies as well as volume. The work done for each interval is its average pressure times the change in volume, or the area under the curve over that interval. Thus the total area under the curve equals the total work done. (b) Work must be done on the system to follow the reverse path. This is interpreted as a negative area under the curve.

We can see where this leads by considering [link](a), which shows a more general process in which both pressure and volume change. The area under the curve is closely approximated by dividing it into strips, each having an average constant pressure $P_{i(\text{ave})}$. The work done is $W_i = P_{i(\text{ave})}\Delta V_i$ for each strip, and the total work done is the sum of the W_i. Thus the total work done is the total area under the curve. If the path is reversed, as in [link](b), then work is done on the system. The area under the curve in that case is negative, because ΔV is negative.

PV diagrams clearly illustrate that *the work done depends on the path taken and not just the endpoints*. This path dependence is seen in [link](a), where more work is done in going from A to C by the path via point B than by the path via point D. The vertical paths, where volume is constant, are called **isochoric** processes. Since volume is constant, $\Delta V=0$, and no work is done in an isochoric process. Now, if the system follows the cyclical path ABCDA, as in [link](b), then the total work done is the area inside the loop. The negative area below path CD subtracts, leaving only the area inside the rectangle. In fact, the work done in any cyclical process (one that returns to its starting point) is the area inside the loop it forms on a PV diagram, as [link](c) illustrates for a general cyclical process. Note that the loop must be traversed in the clockwise direction for work to be positive—that is, for there to be a net work output.


P
Isobaric
$W_{ABC} > W_{ADC}$
A
B
D
C
Isochoric
V
(a)
P_{AB}
P
1.5×10^6 N/m^2
A
B
P_{CD}
2×10^5 N/m^2
$W_{tot} = W_{out} + W_{in}$
D
C
V
$\Delta V = 500$ cm^3
(b)
P
W = area inside loop
V
(c)

(a) The work done in going from A to C depends on path. The work is greater for the path ABC than for the path ADC, because the former is at higher pressure. In both cases, the work done is the area under the path. This area is greater for path ABC. (b) The total work done in the cyclical process

ABCDA is the area inside the loop, since the negative area below CD subtracts out, leaving just the area inside the rectangle. (The values given for the pressures and the change in volume are intended for use in the example below.) (c) The area inside any closed loop is the work done in the cyclical process. If the loop is traversed in a clockwise direction, W is positive—it is work done on the outside environment. If the loop is traveled in a counter-clockwise direction, W is negative—it is work that is done to the system.

Example:

Total Work Done in a Cyclical Process Equals the Area Inside the Closed Loop on a *PV* Diagram

Calculate the total work done in the cyclical process ABCDA shown in [link](b) by the following two methods to verify that work equals the area inside the closed loop on the PV diagram. (Take the data in the figure to be precise to three significant figures.) (a) Calculate the work done along each segment of the path and add these values to get the total work. (b) Calculate the area inside the rectangle ABCDA.

Strategy

To find the work along any path on a PV diagram, you use the fact that work is pressure times change in volume, or $W = P\Delta V$. So in part (a), this value is calculated for each leg of the path around the closed loop.

Solution for (a)

The work along path AB is

Equation:

$$\begin{aligned} W_{\text{AB}} &= P_{\text{AB}}\Delta V_{\text{AB}} \\ &= (1.50\times10^{6}\ \text{N/m}^2)(5.00\times10^{-4}\ \text{m}^3) = 750\ \text{J}. \end{aligned}$$

Since the path BC is isochoric, $\Delta V_{\text{BC}} = 0$, and so $W_{\text{BC}} = 0$. The work along path CD is negative, since ΔV_{CD} is negative (the volume decreases). The work is

Equation:

$$\begin{aligned} W_{\text{CD}} &= P_{\text{CD}}\Delta V_{\text{CD}} \\ &= (2.00\times10^{5}\ \text{N/m}^2)(-5.00\times10^{-4}\ \text{m}^3) = -100\ \text{J}. \end{aligned}$$

Again, since the path DA is isochoric, $\Delta V_{\text{DA}} = 0$, and so $W_{\text{DA}} = 0$. Now the total work is

Equation:

$$\begin{aligned} W &= W_{\text{AB}} + W_{\text{BC}} + W_{\text{CD}} + W_{\text{DA}} \\ &= 750\ \text{J} + 0 + (-100\text{J}) + 0 = 650\ \text{J}. \end{aligned}$$

Solution for (b)

The area inside the rectangle is its height times its width, or

Equation:

$$\begin{aligned} \text{area} &= (P_{\text{AB}} - P_{\text{CD}})\Delta V \\ &= \left[(1.50\times10^{6}\ \text{N/m}^2) - (2.00\times10^{5}\ \text{N/m}^2)\right](5.00\times10^{-4}\ \text{m}^3) \\ &= 650\ \text{J}. \end{aligned}$$

Thus,

Equation:

$$\text{area} = 650\ \text{J} = W.$$

Discussion

The result, as anticipated, is that the area inside the closed loop equals the work done. The area is often easier to calculate than is the work done along each path. It is also convenient to visualize the area inside different curves on PV diagrams in order to see which processes might produce the most work. Recall that work can be done to the system, or by the system, depending on the sign of W. A positive W is work that is done by the system on the outside environment; a negative W represents work done by the environment on the system.

[link](a) shows two other important processes on a PV diagram. For comparison, both are shown starting from the same point A. The upper curve ending at point B is an **isothermal** process—that is, one in which temperature is kept constant. If the gas behaves like an ideal gas, as is often the case, and if no phase change occurs, then $\mathrm{PV}=\mathrm{nRT}$. Since T is constant, PV is a constant for an isothermal process. We ordinarily expect the temperature of a gas to decrease as it expands, and so we correctly suspect that heat transfer must occur from the surroundings to the gas to keep the temperature constant during an isothermal expansion. To show this more rigorously for the special case of a monatomic ideal gas, we note that the average kinetic energy of an atom in such a gas is given by

Equation:

$$\frac{1}{2}mv^2=\frac{3}{2}\mathrm{kT}.$$

The kinetic energy of the atoms in a monatomic ideal gas is its only form of internal energy, and so its total internal energy U is

Equation:

$$U=N\frac{1}{2}mv^2=\frac{3}{2}\mathrm{NkT},\ \text{(monatomic ideal gas)},$$

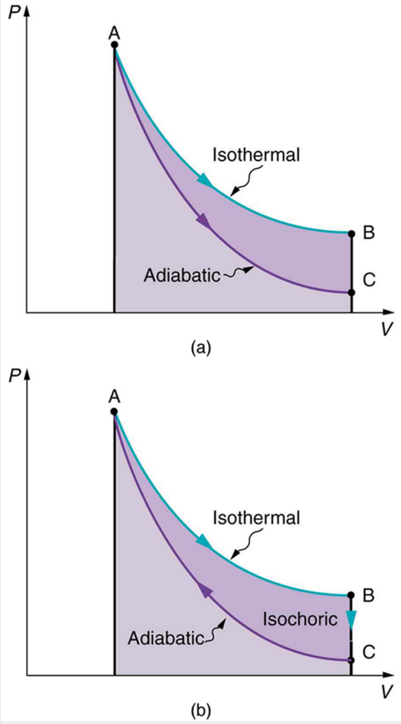
where N is the number of atoms in the gas. This relationship means that the internal energy of an ideal monatomic gas is constant during an isothermal process—that is, $\Delta U=0$. If the internal energy does not change, then the net heat transfer into the gas must equal the net work done by the gas. That is, because $\Delta U=Q-W=0$ here, $Q=W$. We must have just enough heat transfer to replace the work done. An isothermal process is inherently slow, because heat transfer occurs continuously to keep the gas temperature constant at all times and must be allowed to spread through the gas so that there are no hot or cold regions.

Also shown in [link](a) is a curve AC for an **adiabatic** process, defined to be one in which there is no heat transfer—that is, $Q = 0$. Processes that are nearly adiabatic can be achieved either by using very effective insulation or by performing the process so fast that there is little time for heat transfer. Temperature must decrease during an adiabatic expansion process, since work is done at the expense of internal energy:

Equation:

$$U = \frac{3}{2}\text{NkT}.$$

(You might have noted that a gas released into atmospheric pressure from a pressurized cylinder is substantially colder than the gas in the cylinder.) In fact, because $Q = 0$, $\Delta U = -W$ for an adiabatic process. Lower temperature results in lower pressure along the way, so that curve AC is lower than curve AB, and less work is done. If the path ABCA could be followed by cooling the gas from B to C at constant volume (isochorically), [link](b), there would be a net work output.



(a) The upper curve is an isothermal process ($\Delta T = 0$), whereas the lower curve is an adiabatic process ($Q = 0$). Both start from the same point A, but the isothermal process does more work than the adiabatic because heat transfer into the gas takes place to keep its temperature constant. This keeps the pressure higher all along the isothermal path than along the adiabatic path, producing more work. The adiabatic path thus ends up with a lower pressure and temperature at point C, even though the final volume is the same as for the isothermal process. (b) The cycle ABCA produces a net work output.

Reversible Processes

Both isothermal and adiabatic processes such as shown in [link] are reversible in principle. A **reversible process** is one in which both the system and its environment can return to exactly the states they were in by following the reverse path. The reverse isothermal and adiabatic paths are BA and CA, respectively. Real macroscopic processes are never exactly reversible. In the previous examples, our system is a gas (like that in [link]), and its environment is the piston, cylinder, and the rest of the universe. If there are any energy-dissipating mechanisms, such as friction or turbulence, then heat transfer to the environment occurs for either direction of the piston. So, for example, if the path BA is followed and there is friction, then the gas will be returned to its original state but the environment will not—it will have been heated in both directions. Reversibility requires the direction of heat transfer to reverse for the reverse path. Since dissipative mechanisms cannot be completely eliminated, real processes cannot be reversible.

There must be reasons that real macroscopic processes cannot be reversible. We can imagine them going in reverse. For example, heat transfer occurs spontaneously from hot to cold and never spontaneously the reverse. Yet it would not violate the first law of thermodynamics for this to happen. In fact, all spontaneous processes, such as bubbles bursting, never go in reverse. There is a second thermodynamic law that forbids them from going in reverse. When we study this law, we will learn something about nature and also find that such a law limits the efficiency of heat engines. We will find that heat engines with the greatest possible theoretical efficiency would have to use reversible processes, and even they cannot convert all heat transfer into doing work. [link] summarizes the simpler thermodynamic processes and their definitions.

Isobaric	Constant pressure $$W = P\Delta V$$
Isochoric	Constant volume $$W = 0$$
Isothermal	Constant temperature $$Q = W$$
Adiabatic	No heat transfer $$Q = 0$$

Summary of Simple Thermodynamic Processes

Note:
PhET Explorations: States of Matter
Watch different types of molecules form a solid, liquid, or gas. Add or remove heat and watch the phase change. Change the temperature or volume of a container and see a pressure-temperature diagram respond in real time. Relate the interaction potential to the forces between molecules.
https://phet.colorado.edu/sims/html/states-of-matter/latest/states-of-matter_en.html

Section Summary

- One of the important implications of the first law of thermodynamics is that machines can be harnessed to do work that humans previously did by hand or by external energy supplies such as running water or the heat of the Sun. A machine that uses heat transfer to do work is known as a heat engine.

- There are several simple processes, used by heat engines, that flow from the first law of thermodynamics. Among them are the isobaric, isochoric, isothermal and adiabatic processes.
- These processes differ from one another based on how they affect pressure, volume, temperature, and heat transfer.
- If the work done is performed on the outside environment, work (W) will be a positive value. If the work done is done to the heat engine system, work (W) will be a negative value.
- Some thermodynamic processes, including isothermal and adiabatic processes, are reversible in theory; that is, both the thermodynamic system and the environment can be returned to their initial states. However, because of loss of energy owing to the second law of thermodynamics, complete reversibility does not work in practice.

Conceptual Questions

Exercise:

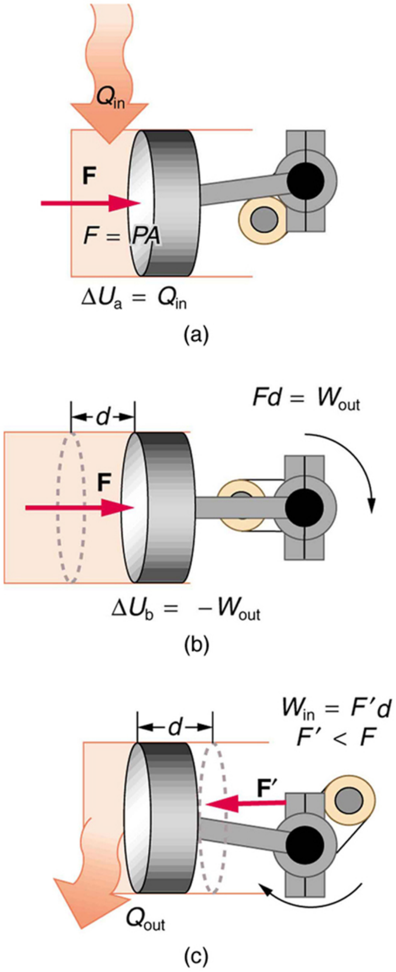
Problem:

A great deal of effort, time, and money has been spent in the quest for the so-called perpetual-motion machine, which is defined as a hypothetical machine that operates or produces useful work indefinitely and/or a hypothetical machine that produces more work or energy than it consumes. Explain, in terms of heat engines and the first law of thermodynamics, why or why not such a machine is likely to be constructed.

Exercise:

Problem:

One method of converting heat transfer into doing work is for heat transfer into a gas to take place, which expands, doing work on a piston, as shown in the figure below. (a) Is the heat transfer converted directly to work in an isobaric process, or does it go through another form first? Explain your answer. (b) What about in an isothermal process? (c) What about in an adiabatic process (where heat transfer occurred prior to the adiabatic process)?



(a)

(b)

(c)

Exercise:

Problem:

Would the previous question make any sense for an isochoric process? Explain your answer.

Exercise:

Problem:

We ordinarily say that $\Delta U = 0$ for an isothermal process. Does this assume no phase change takes place? Explain your answer.

Exercise:

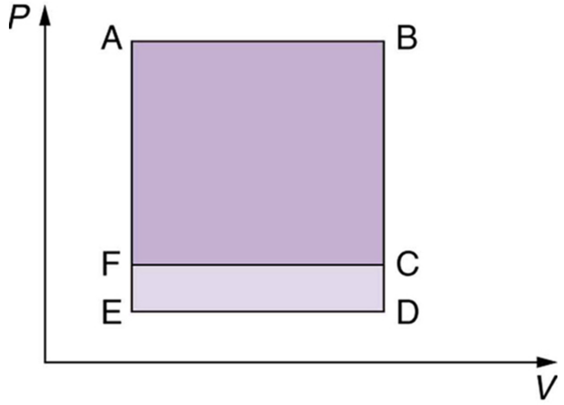
Problem:

The temperature of a rapidly expanding gas decreases. Explain why in terms of the first law of thermodynamics. (Hint: Consider whether the gas does work and whether heat transfer occurs rapidly into the gas through conduction.)

Exercise:

Problem:

Which cyclical process represented by the two closed loops, ABCFA and ABDEA, on the PV diagram in the figure below produces the greatest *net* work? Is that process also the one with the smallest work input required to return it to point A? Explain your responses.



The two cyclical processes shown on this PV diagram start with and return the system to the conditions at point A, but they follow

different paths and produce
different amounts of work.

Exercise:

Problem:

A real process may be nearly adiabatic if it occurs over a very short time. How does the short time span help the process to be adiabatic?

Exercise:

Problem:

It is unlikely that a process can be isothermal unless it is a very slow process. Explain why. Is the same true for isobaric and isochoric processes? Explain your answer.

Problem Exercises

Exercise:

Problem:

A car tire contains $0.0380\ \text{m}^3$ of air at a pressure of $2.20\times10^5\ \text{N/m}^2$ (about 32 psi). How much more internal energy does this gas have than the same volume has at zero gauge pressure (which is equivalent to normal atmospheric pressure)?

Solution:

$6.77\times10^3\ \text{J}$

Exercise:

Problem:

A helium-filled toy balloon has a gauge pressure of 0.200 atm and a volume of 10.0 L. How much greater is the internal energy of the helium in the balloon than it would be at zero gauge pressure?

Exercise:

Problem:

Steam to drive an old-fashioned steam locomotive is supplied at a constant gauge pressure of $1.75 \times 10^6\ \text{N/m}^2$ (about 250 psi) to a piston with a 0.200-m radius. (a) By calculating $P\Delta V$, find the work done by the steam when the piston moves 0.800 m. Note that this is the net work output, since gauge pressure is used. (b) Now find the amount of work by calculating the force exerted times the distance traveled. Is the answer the same as in part (a)?

Solution:

(a) $W = P\Delta V = 1.76 \times 10^5\ \text{J}$

(b) $W = \text{Fd} = 1.76 \times 10^5\ \text{J}$. Yes, the answer is the same.

Exercise:

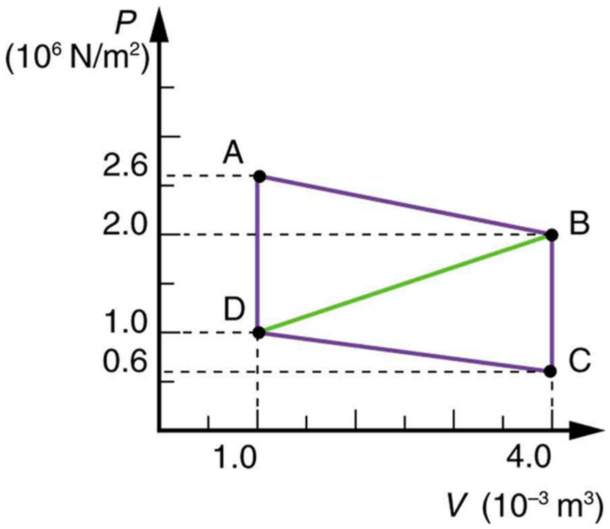
Problem:

A hand-driven tire pump has a piston with a 2.50-cm diameter and a maximum stroke of 30.0 cm. (a) How much work do you do in one stroke if the average gauge pressure is $2.40 \times 10^5\ \text{N/m}^2$ (about 35 psi)? (b) What average force do you exert on the piston, neglecting friction and gravitational force?

Exercise:

Problem:

Calculate the net work output of a heat engine following path ABCDA in the figure below.



Solution:

$W = 4.5 \times 10^3$ J

Exercise:

Problem:

What is the net work output of a heat engine that follows path ABDA in the figure above, with a straight line from B to D? Why is the work output less than for path ABCDA? Explicitly show how you follow the steps in the Problem-Solving Strategies for Thermodynamics.

Exercise:

Problem: **Unreasonable Results**

What is wrong with the claim that a cyclical heat engine does 4.00 kJ of work on an input of 24.0 kJ of heat transfer while 16.0 kJ of heat transfers to the environment?

Solution:

W is not equal to the difference between the heat input and the heat output.

Exercise:

Problem:

(a) A cyclical heat engine, operating between temperatures of 450^o C and 150^o C produces 4.00 MJ of work on a heat transfer of 5.00 MJ into the engine. How much heat transfer occurs to the environment? (b) What is unreasonable about the engine? (c) Which premise is unreasonable?

Exercise:

Problem: Construct Your Own Problem

Consider a car’s gasoline engine. Construct a problem in which you calculate the maximum efficiency this engine can have. Among the things to consider are the effective hot and cold reservoir temperatures. Compare your calculated efficiency with the actual efficiency of car engines.

Exercise:

Problem: Construct Your Own Problem

Consider a car trip into the mountains. Construct a problem in which you calculate the overall efficiency of the car for the trip as a ratio of kinetic and potential energy gained to fuel consumed. Compare this efficiency to the thermodynamic efficiency quoted for gasoline engines and discuss why the thermodynamic efficiency is so much greater. Among the factors to be considered are the gain in altitude and speed, the mass of the car, the distance traveled, and typical fuel economy.

Glossary

heat engine	a machine that uses heat transfer to do work
isobaric process	constant-pressure process in which a gas does work

isochoric process
a constant-volume process

isothermal process
a constant-temperature process

adiabatic process
a process in which no heat transfer takes place

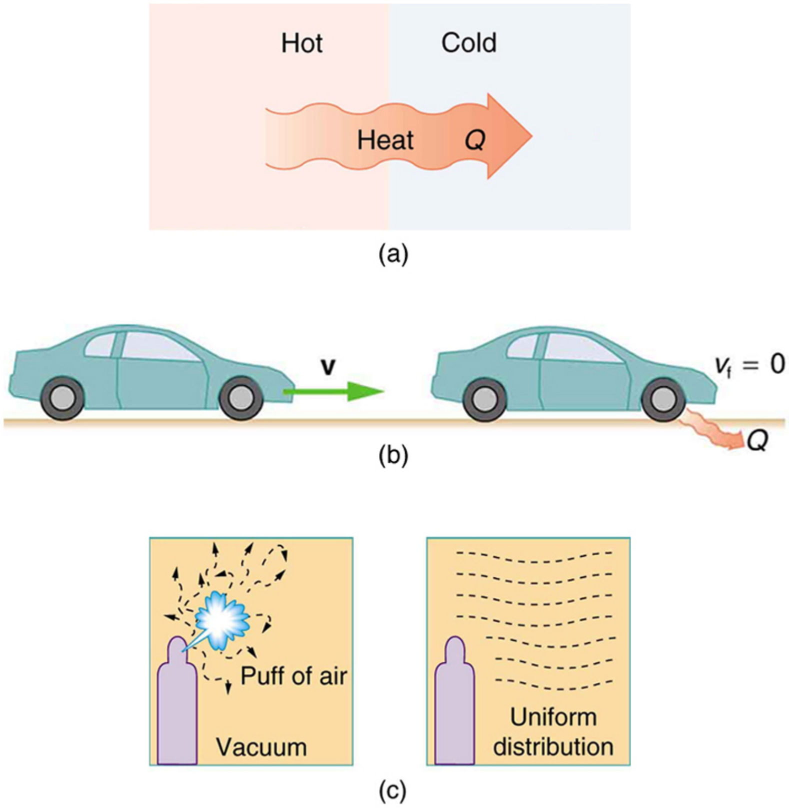
reversible process
a process in which both the heat engine system and the external environment theoretically can be returned to their original states

Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency

- State the expressions of the second law of thermodynamics.
- Calculate the efficiency and carbon dioxide emission of a coal-fired electricity plant, using second law characteristics.
- Describe and define the Otto cycle.

These ice floes melt during the Arctic summer. Some of them refreeze in the winter, but the second law of thermodynamics predicts that it would be extremely unlikely for the water molecules contained in these particular floes to reform the distinctive alligator-like shape they formed when the picture was taken in the summer of 2009. (credit: Patrick Kelley, U.S. Coast Guard, U.S. Geological Survey)

The second law of thermodynamics deals with the direction taken by spontaneous processes. Many processes occur spontaneously in one direction only—that is, they are irreversible, under a given set of conditions. Although irreversibility is seen in day-to-day life—a broken glass does not resume its original state, for instance—complete irreversibility is a statistical statement that cannot be seen during the lifetime of the universe. More precisely, an **irreversible process** is one that depends on path. If the process can go in only one direction, then the reverse path differs fundamentally and the process cannot be reversible. For example, as noted in the previous section, heat involves the transfer of energy from higher to lower temperature. A cold object in contact with a hot one never gets colder, transferring heat to the hot object and making it hotter. Furthermore, mechanical energy, such as kinetic energy, can be completely converted to thermal energy by friction, but the reverse is impossible. A hot stationary object never spontaneously cools off and starts moving. Yet another example is the expansion of a puff of gas introduced into one corner of a vacuum chamber. The gas expands to fill the chamber, but it never regroups in the corner. The random motion of the gas molecules could take them all back to the corner, but this is never observed to happen. (See [link].)



Examples of one-way processes in nature.

(a) Heat transfer occurs spontaneously from hot to cold and not from cold to hot. (b) The brakes of this car convert its kinetic energy to heat transfer to the environment. The reverse process is impossible. (c) The burst of gas let into this vacuum chamber quickly expands to uniformly fill every part of the chamber. The random motions of the gas molecules will never return them to the corner.

The fact that certain processes never occur suggests that there is a law forbidding them to occur. The first law of thermodynamics would allow them to occur—none of those processes violate conservation of energy. The law that forbids these processes is called the second law of thermodynamics. We shall see that the second law can be stated in many ways that may seem different, but which in fact are equivalent. Like all natural laws, the second law of thermodynamics gives insights into nature, and its several statements imply that it is broadly applicable, fundamentally affecting many apparently disparate processes.

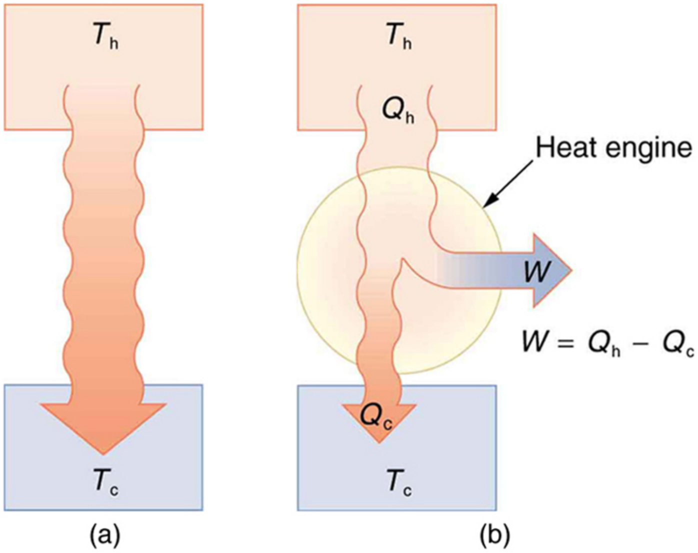
The already familiar direction of heat transfer from hot to cold is the basis of our first version of the **second law of thermodynamics**.

Note:
The Second Law of Thermodynamics (first expression)
Heat transfer occurs spontaneously from higher- to lower-temperature bodies but never spontaneously in the reverse direction.

Another way of stating this: It is impossible for any process to have as its sole result heat transfer from a cooler to a hotter object.

Heat Engines

Now let us consider a device that uses heat transfer to do work. As noted in the previous section, such a device is called a heat engine, and one is shown schematically in [link](b). Gasoline and diesel engines, jet engines, and steam turbines are all heat engines that do work by using part of the heat transfer from some source. Heat transfer from the hot object (or hot reservoir) is denoted as Q_{h}, while heat transfer into the cold object (or cold reservoir) is Q_{c}, and the work done by the engine is W. The temperatures of the hot and cold reservoirs are T_{h} and T_{c}, respectively.



(a) Heat transfer occurs spontaneously from a hot object to a cold one, consistent with the second law of thermodynamics. (b) A heat engine, represented here by a circle, uses part of the heat transfer to do work. The hot and cold objects are called the hot and cold reservoirs. Q_{h} is the heat transfer out of the hot reservoir, W is the work output, and Q_{c} is the heat transfer into the cold reservoir.

Because the hot reservoir is heated externally, which is energy intensive, it is important that the work is done as efficiently as possible. In fact, we would like W to equal Q_{h}, and for there to be no heat transfer to the environment ($Q_{\mathrm{c}} = 0$). Unfortunately, this is impossible. The **second law of thermodynamics** also states, with regard to using heat transfer to do work (the second expression of the second law):

Note:
The Second Law of Thermodynamics (second expression)
It is impossible in any system for heat transfer from a reservoir to completely convert to work in a cyclical process in which the system returns to its initial state.

A **cyclical process** brings a system, such as the gas in a cylinder, back to its original state at the end of every cycle. Most heat engines, such as reciprocating piston engines and rotating turbines, use cyclical processes. The second law, just stated in its second form, clearly states that such engines cannot have perfect conversion of heat transfer into work done. Before going into the underlying reasons for the limits on converting heat transfer into work, we need to explore the relationships among W, Q_{h}, and Q_{c}, and to define the efficiency of a cyclical heat engine. As noted, a cyclical process brings the system back to its original condition at the end of every cycle. Such a system's internal energy U is the same at the beginning and end of every cycle—that is, $\Delta U = 0$. The first law of thermodynamics states that

Equation:

$$\Delta U = Q - W,$$

where Q is the *net* heat transfer during the cycle ($Q = Q_{\mathrm{h}} - Q_{\mathrm{c}}$) and W is the net work done by the system. Since $\Delta U = 0$ for a complete cycle, we

have
Equation:

$$0 = Q - W,$$

so that
Equation:

$$W = Q.$$

Thus the net work done by the system equals the net heat transfer into the system, or
Equation:

$$W = Q_{\mathrm{h}} - Q_{\mathrm{c}} \ (\text{cyclical process}),$$

just as shown schematically in [link](b). The problem is that in all processes, there is some heat transfer Q_{c} to the environment—and usually a very significant amount at that.

In the conversion of energy to work, we are always faced with the problem of getting less out than we put in. We define *conversion efficiency* Eff to be the ratio of useful work output to the energy input (or, in other words, the ratio of what we get to what we spend). In that spirit, we define the efficiency of a heat engine to be its net work output W divided by heat transfer to the engine Q_{h}; that is,
Equation:

$$\mathrm{Eff} = \frac{W}{Q_{\mathrm{h}}}.$$

Since $W = Q_{\mathrm{h}} - Q_{\mathrm{c}}$ in a cyclical process, we can also express this as
Equation:

$$\text{Eff} = \frac{Q_\text{h} - Q_\text{c}}{Q_\text{h}} = 1 - \frac{Q_\text{c}}{Q_\text{h}} \text{ (cyclical process)},$$

making it clear that an efficiency of 1, or 100%, is possible only if there is no heat transfer to the environment ($Q_\text{c} = 0$). Note that all Qs are positive. The direction of heat transfer is indicated by a plus or minus sign. For example, Q_c is out of the system and so is preceded by a minus sign.

Example:

Daily Work Done by a Coal-Fired Power Station, Its Efficiency and Carbon Dioxide Emissions

A coal-fired power station is a huge heat engine. It uses heat transfer from burning coal to do work to turn turbines, which are used to generate electricity. In a single day, a large coal power station has $2.50 \times 10^{14}\ \text{J}$ of heat transfer from coal and $1.48 \times 10^{14}\ \text{J}$ of heat transfer into the environment. (a) What is the work done by the power station? (b) What is the efficiency of the power station? (c) In the combustion process, the following chemical reaction occurs: $\text{C} + \text{O}_2 \rightarrow \text{CO}_2$. This implies that every 12 kg of coal puts 12 kg + 16 kg + 16 kg = 44 kg of carbon dioxide into the atmosphere. Assuming that 1 kg of coal can provide $2.5 \times 10^6\ \text{J}$ of heat transfer upon combustion, how much CO_2 is emitted per day by this power plant?

Strategy for (a)

We can use $W = Q_\text{h} - Q_\text{c}$ to find the work output W, assuming a cyclical process is used in the power station. In this process, water is boiled under pressure to form high-temperature steam, which is used to run steam turbine-generators, and then condensed back to water to start the cycle again.

Solution for (a)

Work output is given by:

Equation:

$$W = Q_\text{h} - Q_\text{c}.$$

Substituting the given values:

Equation:

$$\begin{aligned} W &= 2.50\times10^{14}\text{ J}-1.48\times10^{14}\text{ J} \\ &= 1.02\times10^{14}\text{ J}. \end{aligned}$$

Strategy for (b)
The efficiency can be calculated with $\text{Eff}=\frac{W}{Q_{\text{h}}}$ since Q_{h} is given and work W was found in the first part of this example.
Solution for (b)
Efficiency is given by: $\text{Eff}=\frac{W}{Q_{\text{h}}}$. The work W was just found to be 1.02×10^{14} J, and Q_{h} is given, so the efficiency is
Equation:

$$\begin{aligned} Eff &= \frac{1.02\times10^{14}\text{ J}}{2.50\times10^{14}\text{ J}} \\ &= 0.408\text{, or }40.8\% \end{aligned}$$

Strategy for (c)
The daily consumption of coal is calculated using the information that each day there is 2.50×10^{14} J of heat transfer from coal. In the combustion process, we have $\text{C}+\text{O}_2\rightarrow\text{CO}_2$. So every 12 kg of coal puts 12 kg + 16 kg + 16 kg = 44 kg of CO_2 into the atmosphere.
Solution for (c)
The daily coal consumption is
Equation:

$$\frac{2.50\times10^{14}\text{ J}}{2.50\times10^{6}\text{ J/kg}}=1.0\times10^{8}\text{ kg}.$$

Assuming that the coal is pure and that all the coal goes toward producing carbon dioxide, the carbon dioxide produced per day is
Equation:

$$1.0\times10^{8}\text{ kg coal}\times\frac{44\text{ kg CO}_2}{12\text{ kg coal}}=3.7\times10^{8}\text{ kg CO}_2.$$

This is 370,000 metric tons of CO_2 produced every day.

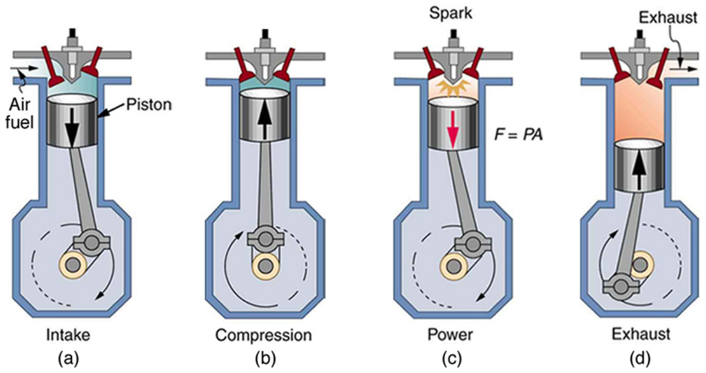
Discussion
If all the work output is converted to electricity in a period of one day, the average power output is 1180 MW (this is left to you as an end-of-chapter problem). This value is about the size of a large-scale conventional power plant. The efficiency found is acceptably close to the value of 42% given for coal power stations. It means that fully 59.2% of the energy is heat transfer to the environment, which usually results in warming lakes, rivers, or the ocean near the power station, and is implicated in a warming planet generally. While the laws of thermodynamics limit the efficiency of such plants—including plants fired by nuclear fuel, oil, and natural gas—the heat transfer to the environment could be, and sometimes is, used for heating homes or for industrial processes. The generally low cost of energy has not made it economical to make better use of the waste heat transfer from most heat engines. Coal-fired power plants produce the greatest amount of CO_2 per unit energy output (compared to natural gas or oil), making coal the least efficient fossil fuel.

With the information given in [link], we can find characteristics such as the efficiency of a heat engine without any knowledge of how the heat engine operates, but looking further into the mechanism of the engine will give us greater insight. [link] illustrates the operation of the common four-stroke gasoline engine. The four steps shown complete this heat engine’s cycle, bringing the gasoline-air mixture back to its original condition.

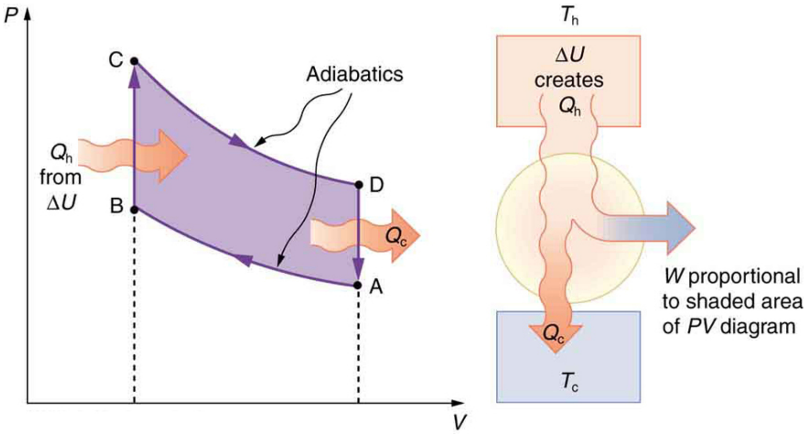
The **Otto cycle** shown in [link](a) is used in four-stroke internal combustion engines, although in fact the true Otto cycle paths do not correspond exactly to the strokes of the engine.

The adiabatic process AB corresponds to the nearly adiabatic compression stroke of the gasoline engine. In both cases, work is done on the system (the gas mixture in the cylinder), increasing its temperature and pressure. Along path BC of the Otto cycle, heat transfer Q_h into the gas occurs at constant volume, causing a further increase in pressure and temperature. This process corresponds to burning fuel in an internal combustion engine, and takes place so rapidly that the volume is nearly constant. Path CD in the Otto cycle is an adiabatic expansion that does work on the outside world, just as the power stroke of an internal combustion engine does in its nearly adiabatic expansion. The work done by the system along path CD is greater than the work done on the system along path AB, because the pressure is greater, and so there is a net work output. Along path DA in the Otto cycle, heat transfer Q_c from the gas at constant volume reduces its temperature and pressure, returning it to its original state. In an internal combustion engine, this process corresponds to the exhaust of hot gases and the intake of an air-gasoline mixture at a considerably lower temperature. In both cases, heat transfer into the environment occurs along this final path.

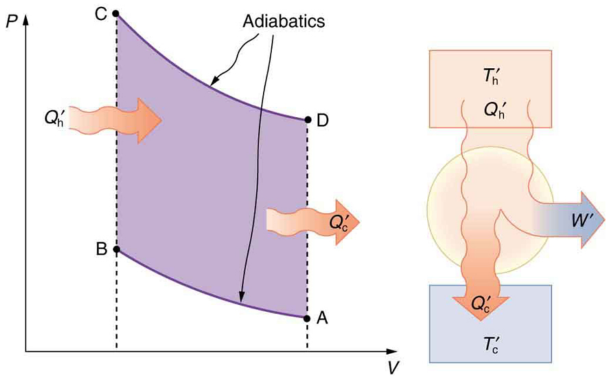
The net work done by a cyclical process is the area inside the closed path on a PV diagram, such as that inside path ABCDA in [link]. Note that in every imaginable cyclical process, it is absolutely necessary for heat transfer from the system to occur in order to get a net work output. In the Otto cycle, heat transfer occurs along path DA. If no heat transfer occurs, then the return path is the same, and the net work output is zero. The lower the temperature on the path AB, the less work has to be done to compress the gas. The area inside the closed path is then greater, and so the engine does more work and is thus more efficient. Similarly, the higher the temperature along path CD, the more work output there is. (See [link].) So efficiency is related to the temperatures of the hot and cold reservoirs. In the next section, we shall see what the absolute limit to the efficiency of a heat engine is, and how it is related to temperature.



In the four-stroke internal combustion gasoline engine, heat transfer into work takes place in the cyclical process shown here. The piston is connected to a rotating crankshaft, which both takes work out of and does work on the gas in the cylinder. (a) Air is mixed with fuel during the intake stroke. (b) During the compression stroke, the air-fuel mixture is rapidly compressed in a nearly adiabatic process, as the piston rises with the valves closed. Work is done on the gas. (c) The power stroke has two distinct parts. First, the air-fuel mixture is ignited, converting chemical potential energy into thermal energy almost instantaneously, which leads to a great increase in pressure. Then the piston descends, and the gas does work by exerting a force through a distance in a nearly adiabatic process. (d) The exhaust stroke expels the hot gas to prepare the engine for another cycle, starting again with the intake stroke.


P
V
C
B
D
A
Adiabatics
Q_h
from
ΔU
Q_c
T_h
ΔU
creates
Q_h
Q_c
T_c
W proportional
to shaded area
of PV diagram

PV diagram for a simplified Otto cycle, analogous to that employed in an internal combustion engine. Point A corresponds to the start of the compression stroke of an internal combustion engine. Paths AB and CD are adiabatic and correspond to the compression and power strokes of an internal combustion engine, respectively. Paths BC and DA are isochoric and accomplish similar results to the ignition and exhaust-intake portions, respectively, of the internal combustion engine's cycle. Work is done on the gas along path AB, but more work is done by the gas along path CD, so that there is a net work output.



This Otto cycle produces a greater work output than the one in [link], because the starting temperature of path CD is higher and the starting temperature of path AB is lower. The area inside the loop is greater, corresponding to greater net work output.

Section Summary

- The two expressions of the second law of thermodynamics are: (i) Heat transfer occurs spontaneously from higher- to lower-temperature bodies but never spontaneously in the reverse direction; and (ii) It is impossible in any system for heat transfer from a reservoir to completely convert to work in a cyclical process in which the system returns to its initial state.
- Irreversible processes depend on path and do not return to their original state. Cyclical processes are processes that return to their original state at the end of every cycle.
- In a cyclical process, such as a heat engine, the net work done by the system equals the net heat transfer into the system, or $W = Q_{\text{h}} - Q_{\text{c}}$, where Q_{h} is the heat transfer from the hot object (hot reservoir), and Q_{c} is the heat transfer into the cold object (cold reservoir).

- Efficiency can be expressed as $\text{Eff} = \frac{W}{Q_{\text{h}}}$, the ratio of work output divided by the amount of energy input.
- The four-stroke gasoline engine is often explained in terms of the Otto cycle, which is a repeating sequence of processes that convert heat into work.

Conceptual Questions

Exercise:

Problem:

Imagine you are driving a car up Pike’s Peak in Colorado. To raise a car weighing 1000 kilograms a distance of 100 meters would require about a million joules. You could raise a car 12.5 kilometers with the energy in a gallon of gas. Driving up Pike's Peak (a mere 3000-meter climb) should consume a little less than a quart of gas. But other considerations have to be taken into account. Explain, in terms of efficiency, what factors may keep you from realizing your ideal energy use on this trip.

Exercise:

Problem:

Is a temperature difference necessary to operate a heat engine? State why or why not.

Exercise:

Problem:

Definitions of efficiency vary depending on how energy is being converted. Compare the definitions of efficiency for the human body and heat engines. How does the definition of efficiency in each relate to the type of energy being converted into doing work?

Exercise:

Problem:

Why—other than the fact that the second law of thermodynamics says reversible engines are the most efficient—should heat engines employing reversible processes be more efficient than those employing irreversible processes? Consider that dissipative mechanisms are one cause of irreversibility.

Problem Exercises

Exercise:

Problem:

A certain heat engine does 10.0 kJ of work and 8.50 kJ of heat transfer occurs to the environment in a cyclical process. (a) What was the heat transfer into this engine? (b) What was the engine's efficiency?

Solution:

(a) 18.5 kJ

(b) 54.1%

Exercise:

Problem:

With $2.56\times10^{6}\text{ J}$ of heat transfer into this engine, a given cyclical heat engine can do only $1.50\times10^{5}\text{ J}$ of work. (a) What is the engine's efficiency? (b) How much heat transfer to the environment takes place?

Exercise:

Problem:

(a) What is the work output of a cyclical heat engine having a 22.0% efficiency and 6.00×10^9 J of heat transfer into the engine? (b) How much heat transfer occurs to the environment?

Solution:

(a) 1.32×10^9 J

(b) 4.68×10^9 J

Exercise:

Problem:

(a) What is the efficiency of a cyclical heat engine in which 75.0 kJ of heat transfer occurs to the environment for every 95.0 kJ of heat transfer into the engine? (b) How much work does it produce for 100 kJ of heat transfer into the engine?

Exercise:

Problem:

The engine of a large ship does 2.00×10^8 J of work with an efficiency of 5.00%. (a) How much heat transfer occurs to the environment? (b) How many barrels of fuel are consumed, if each barrel produces 6.00×10^9 J of heat transfer when burned?

Solution:

(a) 3.80×10^9 J

(b) 0.667 barrels

Exercise:

Problem:

(a) How much heat transfer occurs to the environment by an electrical power station that uses $1.25 \times 10^{14}\ \text{J}$ of heat transfer into the engine with an efficiency of 42.0%? (b) What is the ratio of heat transfer to the environment to work output? (c) How much work is done?

Exercise:

Problem:

Assume that the turbines at a coal-powered power plant were upgraded, resulting in an improvement in efficiency of 3.32%. Assume that prior to the upgrade the power station had an efficiency of 36% and that the heat transfer into the engine in one day is still the same at $2.50 \times 10^{14}\ \text{J}$. (a) How much more electrical energy is produced due to the upgrade? (b) How much less heat transfer occurs to the environment due to the upgrade?

Solution:

(a) $8.30 \times 10^{12}\ \text{J}$, which is 3.32% of $2.50 \times 10^{14}\ \text{J}$.

(b) $-8.30 \times 10^{12}\ \text{J}$, where the negative sign indicates a reduction in heat transfer to the environment.

Exercise:

Problem:

This problem compares the energy output and heat transfer to the environment by two different types of nuclear power stations—one with the normal efficiency of 34.0%, and another with an improved efficiency of 40.0%. Suppose both have the same heat transfer into the engine in one day, $2.50 \times 10^{14}\ \mathrm{J}$. (a) How much more electrical energy is produced by the more efficient power station? (b) How much less heat transfer occurs to the environment by the more efficient power station? (One type of more efficient nuclear power station, the gas-cooled reactor, has not been reliable enough to be economically feasible in spite of its greater efficiency.)

Glossary

irreversible process
any process that depends on path direction

second law of thermodynamics
heat transfer flows from a hotter to a cooler object, never the reverse, and some heat energy in any process is lost to available work in a cyclical process

cyclical process
a process in which the path returns to its original state at the end of every cycle

Otto cycle
a thermodynamic cycle, consisting of a pair of adiabatic processes and a pair of isochoric processes, that converts heat into work, e.g., the four-stroke engine cycle of intake, compression, ignition, and exhaust

Carnot’s Perfect Heat Engine: The Second Law of Thermodynamics Restated

- Identify a Carnot cycle.
- Calculate maximum theoretical efficiency of a nuclear reactor.
- Explain how dissipative processes affect the ideal Carnot engine.

This novelty toy, known as the drinking bird, is an example of Carnot’s engine. It contains methylene chloride (mixed with a dye) in the abdomen, which boils at a very low temperature—about 100°F. To operate, one gets the bird’s head wet. As the water evaporates, fluid moves up into the head, causing the bird to become top-heavy and dip forward back into the water. This cools down the methylene chloride in the head, and it moves back into the abdomen, causing the bird to become bottom heavy and tip up. Except for a very small input of energy—the original head-wetting—the bird becomes a perpetual motion machine of sorts. (credit: Arabesk.nl, Wikimedia Commons)

We know from the second law of thermodynamics that a heat engine cannot be 100% efficient, since there must always be some heat transfer Q_c to the environment, which is often called waste heat. How efficient, then, can a heat engine be? This question was answered at a theoretical level in 1824 by a young French engineer, Sadi Carnot (1796–1832), in his study of the then-emerging heat engine technology crucial to the Industrial Revolution. He devised a theoretical cycle, now called the **Carnot cycle**, which is the most efficient cyclical process possible. The second law of thermodynamics can be restated in terms of the Carnot cycle, and so what Carnot actually discovered was this fundamental law. Any heat engine employing the Carnot cycle is called a **Carnot engine**.

What is crucial to the Carnot cycle—and, in fact, defines it—is that only reversible processes are used. Irreversible processes involve dissipative factors, such as friction and turbulence. This increases heat transfer Q_c to the environment and reduces the efficiency of the engine. Obviously, then, reversible processes are superior.

Note:
Carnot Engine
Stated in terms of reversible processes, the **second law of thermodynamics** has a third form:
A Carnot engine operating between two given temperatures has the greatest possible efficiency of any heat engine operating between these two temperatures. Furthermore, all engines employing only reversible processes have this same maximum efficiency when operating between the same given temperatures.

[link] shows the PV diagram for a Carnot cycle. The cycle comprises two isothermal and two adiabatic processes. Recall that both isothermal and adiabatic processes are, in principle, reversible.

Carnot also determined the efficiency of a perfect heat engine—that is, a Carnot engine. It is always true that the efficiency of a cyclical heat engine is given by:

Equation:

$$\text{Eff} = \frac{Q_\text{h} - Q_\text{c}}{Q_\text{h}} = 1 - \frac{Q_\text{c}}{Q_\text{h}}.$$

What Carnot found was that for a perfect heat engine, the ratio Q_c/Q_h equals the ratio of the absolute temperatures of the heat reservoirs. That is, $Q_\text{c}/Q_\text{h} = T_\text{c}/T_\text{h}$ for a Carnot engine, so that the maximum or **Carnot efficiency** Eff_C is given by

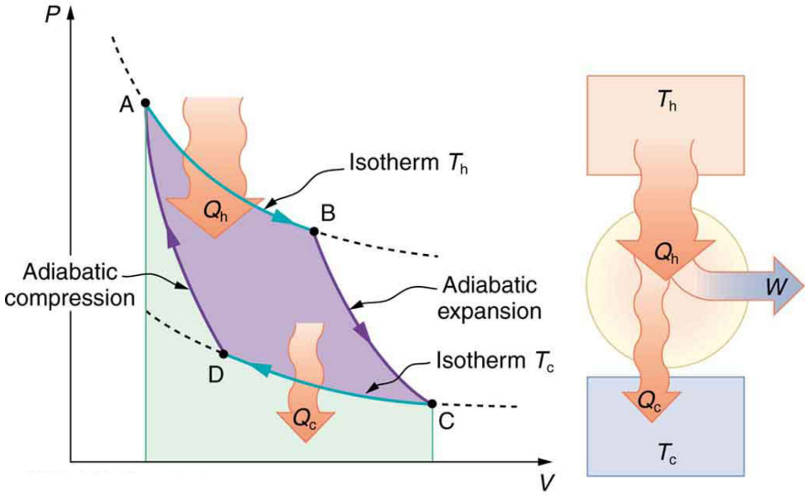
Equation:

$$Eff_\text{C} = 1 - \frac{T_\text{c}}{T_\text{h}},$$

where T_h and T_c are in kelvins (or any other absolute temperature scale). No real heat engine can do as well as the Carnot efficiency—an actual efficiency of about 0.7 of this maximum is usually the best that can be accomplished. But the ideal Carnot engine, like the drinking bird above, while a fascinating novelty, has zero power. This makes it unrealistic for any applications.

Carnot's interesting result implies that 100% efficiency would be possible only if $T_\text{c} = 0\text{ K}$ —that is, only if the cold reservoir were at absolute zero, a practical and theoretical impossibility. But the physical implication is this —the only way to have all heat transfer go into doing work is to remove *all* thermal energy, and this requires a cold reservoir at absolute zero.

It is also apparent that the greatest efficiencies are obtained when the ratio T_c/T_h is as small as possible. Just as discussed for the Otto cycle in the previous section, this means that efficiency is greatest for the highest possible temperature of the hot reservoir and lowest possible temperature of the cold reservoir. (This setup increases the area inside the closed loop on the PV diagram; also, it seems reasonable that the greater the temperature difference, the easier it is to divert the heat transfer to work.) The actual reservoir temperatures of a heat engine are usually related to the type of heat source and the temperature of the environment into which heat transfer occurs. Consider the following example.

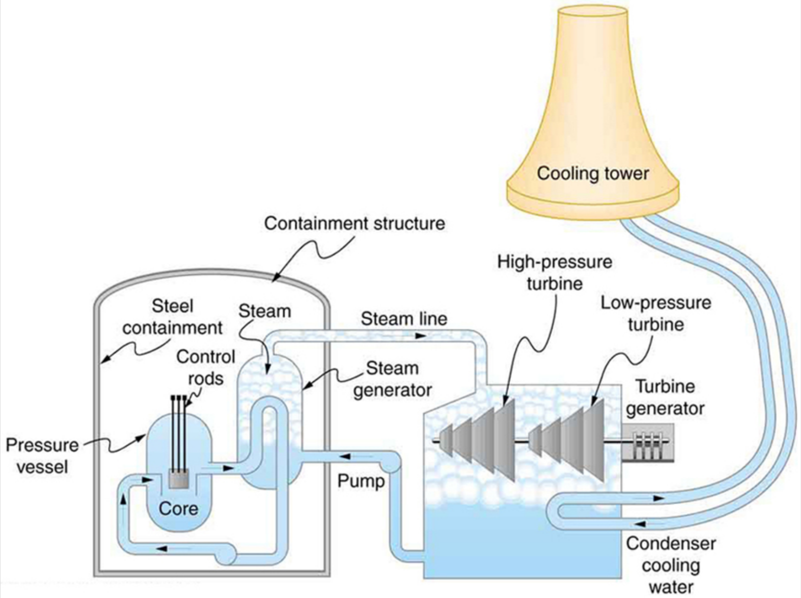

P
A
Isotherm T_h
Q_h
B
Adiabatic compression
Adiabatic expansion
Isotherm T_c
D
Q_c
C
V
T_h
Q_h
W
Q_c
T_c

PV diagram for a Carnot cycle, employing only reversible isothermal and adiabatic processes. Heat transfer Q_h occurs into the working substance during the isothermal path AB, which takes place at constant temperature T_h. Heat transfer Q_c occurs out of the working substance during the isothermal path CD, which takes place at constant temperature T_c. The net work output W equals the area inside the path ABCDA. Also shown is a schematic of a Carnot engine operating between hot and cold reservoirs at temperatures T_h and T_c. Any heat engine using reversible processes and operating between these two temperatures will have the same maximum efficiency as the Carnot engine.

Example:

Maximum Theoretical Efficiency for a Nuclear Reactor

A nuclear power reactor has pressurized water at 300°C. (Higher temperatures are theoretically possible but practically not, due to limitations with materials used in the reactor.) Heat transfer from this water is a complex process (see [link]). Steam, produced in the steam generator, is used to drive the turbine-generators. Eventually the steam is condensed to water at 27°C and then heated again to start the cycle over. Calculate the maximum theoretical efficiency for a heat engine operating between these two temperatures.


Cooling tower
Containment structure
High-pressure turbine
Low-pressure turbine
Steel containment
Steam
Steam line
Control rods
Steam generator
Turbine generator
Pressure vessel
Pump
Core
Condenser cooling water

Schematic diagram of a pressurized water nuclear reactor and the steam turbines that convert work into electrical energy. Heat exchange is used to generate steam, in part to avoid contamination of the generators with radioactivity. Two turbines are used because this is less expensive than operating a single generator that produces the same amount of electrical energy. The steam is condensed to liquid before being returned to the heat exchanger, to keep exit steam pressure low and aid the flow of steam through the turbines (equivalent to using a lower-temperature cold reservoir). The considerable energy associated with condensation must be dissipated into the local environment; in this example, a cooling tower is used so there is no direct heat transfer to an aquatic environment. (Note that the water going to the cooling tower does not come into contact with the steam flowing over the turbines.)

Strategy

Since temperatures are given for the hot and cold reservoirs of this heat engine, $Eff_{\text{C}} = 1 - \frac{T_{\text{c}}}{T_{\text{h}}}$ can be used to calculate the Carnot (maximum theoretical) efficiency. Those temperatures must first be converted to kelvins.

Solution

The hot and cold reservoir temperatures are given as 300^{o}C and 27.0^{o}C, respectively. In kelvins, then, $T_{\text{h}} = 573\ \text{K}$ and $T_{\text{c}} = 300\ \text{K}$, so that the maximum efficiency is

Equation:

$$Eff_{\text{C}} = 1 - \frac{T_{\text{c}}}{T_{\text{h}}}.$$

Thus,

Equation:

$$\begin{aligned} Eff_{\text{C}} &= 1 - \frac{300\ \text{K}}{573\ \text{K}} \\ &= 0.476\text{, or }47.6\%. \end{aligned}$$

Discussion

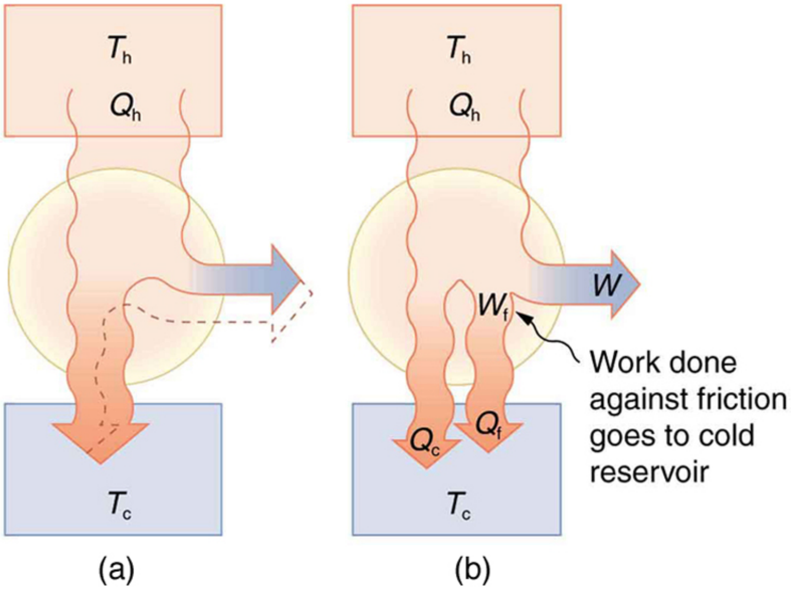
A typical nuclear power station’s actual efficiency is about 35%, a little better than 0.7 times the maximum possible value, a tribute to superior engineering. Electrical power stations fired by coal, oil, and natural gas have greater actual efficiencies (about 42%), because their boilers can reach higher temperatures and pressures. The cold reservoir temperature in any of these power stations is limited by the local environment. [link] shows (a) the exterior of a nuclear power station and (b) the exterior of a coal-fired power station. Both have cooling towers into which water from the condenser enters the tower near the top and is sprayed downward, cooled by evaporation.

(a)

(b)

(a) A nuclear power station (credit: BlatantWorld.com) and (b) a coal-fired power station. Both have cooling towers in which water evaporates into the environment, representing Q_c. The nuclear reactor, which supplies Q_h, is housed inside the dome-shaped containment buildings. (credit: Robert & Mihaela Vicol, publicphoto.org)

Since all real processes are irreversible, the actual efficiency of a heat engine can never be as great as that of a Carnot engine, as illustrated in [link](a). Even with the best heat engine possible, there are always dissipative processes in peripheral equipment, such as electrical transformers or car transmissions. These further reduce the overall efficiency by converting some of the engine's work output back into heat transfer, as shown in [link](b).



Real heat engines are less efficient than Carnot engines. (a) Real engines use irreversible processes, reducing the heat transfer to work. Solid lines represent the actual process; the dashed lines are what a Carnot engine would do between the same two reservoirs. (b) Friction and other dissipative processes in the output mechanisms of a heat engine convert some of its work output into heat transfer to the environment.

Section Summary

- The Carnot cycle is a theoretical cycle that is the most efficient cyclical process possible. Any engine using the Carnot cycle, which uses only reversible processes (adiabatic and isothermal), is known as a Carnot engine.
- Any engine that uses the Carnot cycle enjoys the maximum theoretical efficiency.
- While Carnot engines are ideal engines, in reality, no engine achieves Carnot's theoretical maximum efficiency, since dissipative processes, such as friction, play a role. Carnot cycles without heat loss may be possible at absolute zero, but this has never been seen in nature.

Conceptual Questions

Exercise:

Problem:

Think about the drinking bird at the beginning of this section ([link]). Although the bird enjoys the theoretical maximum efficiency possible, if left to its own devices over time, the bird will cease "drinking." What are some of the dissipative processes that might cause the bird's motion to cease?

Exercise:

Problem:

Can improved engineering and materials be employed in heat engines to reduce heat transfer into the environment? Can they eliminate heat transfer into the environment entirely?

Exercise:

Problem:

Does the second law of thermodynamics alter the conservation of energy principle?

Problem Exercises

Exercise:

Problem:

A certain gasoline engine has an efficiency of 30.0%. What would the hot reservoir temperature be for a Carnot engine having that efficiency, if it operates with a cold reservoir temperature of 200°C?

Solution:

403°C

Exercise:

Problem:

A gas-cooled nuclear reactor operates between hot and cold reservoir temperatures of 700°C and 27.0°C. (a) What is the maximum efficiency of a heat engine operating between these temperatures? (b) Find the ratio of this efficiency to the Carnot efficiency of a standard nuclear reactor (found in [link]).

Exercise:

Problem:

(a) What is the hot reservoir temperature of a Carnot engine that has an efficiency of 42.0% and a cold reservoir temperature of 27.0°C? (b) What must the hot reservoir temperature be for a real heat engine that achieves 0.700 of the maximum efficiency, but still has an efficiency of 42.0% (and a cold reservoir at 27.0°C)? (c) Does your answer imply practical limits to the efficiency of car gasoline engines?

Solution:

(a) 244ºC

(b) 477ºC

(c)Yes, since automobiles engines cannot get too hot without overheating, their efficiency is limited.

Exercise:

Problem:

Steam locomotives have an efficiency of 17.0% and operate with a hot steam temperature of 425ºC. (a) What would the cold reservoir temperature be if this were a Carnot engine? (b) What would the maximum efficiency of this steam engine be if its cold reservoir temperature were 150ºC?

Exercise:

Problem:

Practical steam engines utilize 450ºC steam, which is later exhausted at 270ºC. (a) What is the maximum efficiency that such a heat engine can have? (b) Since 270ºC steam is still quite hot, a second steam engine is sometimes operated using the exhaust of the first. What is the maximum efficiency of the second engine if its exhaust has a temperature of 150ºC? (c) What is the overall efficiency of the two engines? (d) Show that this is the same efficiency as a single Carnot engine operating between 450ºC and 150ºC. Explicitly show how you follow the steps in the Problem-Solving Strategies for Thermodynamics.

Solution:

(a) $Eff_1 = 1 - \frac{T_{\text{c,1}}}{T_{\text{h,1}}} = 1 - \frac{543\text{ K}}{723\text{ K}} = 0.249\text{ or }24.9\%$

(b) $Eff_2 = 1 - \frac{423\text{ K}}{543\text{ K}} = 0.221\text{ or }22.1\%$

(c) $Eff_1 = 1 - \frac{T_{\text{c},1}}{T_{\text{h},1}} \Rightarrow T_{\text{c},1} = T_{\text{h},1}(1, -, eff_1)$
similarly, $T_{\text{c},2} = T_{\text{h},2}(1 - Eff_2)$
using $T_{\text{h},2} = T_{\text{c},1}$ in above equation gives
$T_{\text{c},2} = T_{\text{h},1}(1 - Eff_1)(1 - Eff_2) \equiv T_{\text{h},1}(1 - Eff_{\text{overall}})$
$(1 - Eff_{\text{overall}}) = (1 - Eff_1)(1 - Eff_2)$
$Eff_{\text{overall}} = 1 - (1 - 0.249)(1 - 0.221) = 41.5\%$

(d) $Eff_{\text{overall}} = 1 - \frac{423\text{ K}}{723\text{ K}} = 0.415 \text{ or } 41.5\%$

Exercise:

Problem:

A coal-fired electrical power station has an efficiency of 38%. The temperature of the steam leaving the boiler is 550^{o}C. What percentage of the maximum efficiency does this station obtain? (Assume the temperature of the environment is 20^{o}C.)

Exercise:

Problem:

Would you be willing to financially back an inventor who is marketing a device that she claims has 25 kJ of heat transfer at 600 K, has heat transfer to the environment at 300 K, and does 12 kJ of work? Explain your answer.

Solution:

The heat transfer to the cold reservoir is
$Q_{\text{c}} = Q_{\text{h}} - W = 25\text{ kJ} - 12\text{ kJ} = 13\text{ kJ}$, so the efficiency is
$Eff = 1 - \frac{Q_{\text{c}}}{Q_{\text{h}}} = 1 - \frac{13\text{ kJ}}{25\text{ kJ}} = 0.48$. The Carnot efficiency is
$Eff_{\text{C}} = 1 - \frac{T_{\text{c}}}{T_{\text{h}}} = 1 - \frac{300\text{ K}}{600\text{ K}} = 0.50$. The actual efficiency is 96% of the Carnot efficiency, which is much higher than the best-ever achieved of about 70%, so her scheme is likely to be fraudulent.

Exercise:

Problem: **Unreasonable Results**

(a) Suppose you want to design a steam engine that has heat transfer to the environment at 270°C and has a Carnot efficiency of 0.800. What temperature of hot steam must you use? (b) What is unreasonable about the temperature? (c) Which premise is unreasonable?

Exercise:

Problem: **Unreasonable Results**

Calculate the cold reservoir temperature of a steam engine that uses hot steam at 450°C and has a Carnot efficiency of 0.700. (b) What is unreasonable about the temperature? (c) Which premise is unreasonable?

Solution:

(a) -56.3°C

(b) The temperature is too cold for the output of a steam engine (the local environment). It is below the freezing point of water.

(c) The assumed efficiency is too high.

Glossary

Carnot cycle
: a cyclical process that uses only reversible processes, the adiabatic and isothermal processes

Carnot engine
: a heat engine that uses a Carnot cycle

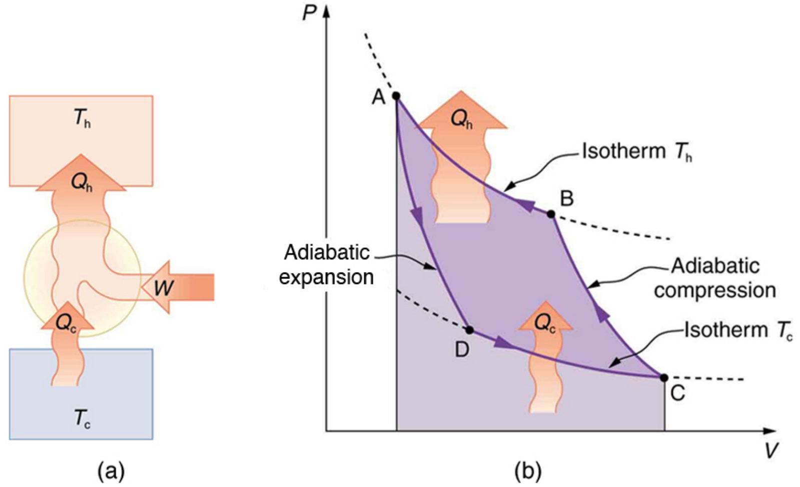
Carnot efficiency
: the maximum theoretical efficiency for a heat engine

Applications of Thermodynamics: Heat Pumps and Refrigerators

- Describe the use of heat engines in heat pumps and refrigerators.
- Demonstrate how a heat pump works to warm an interior space.
- Explain the differences between heat pumps and refrigerators.
- Calculate a heat pump's coefficient of performance.

Almost every home contains a refrigerator. Most people don't realize they are also sharing their homes with a heat pump. (credit: Id1337x, Wikimedia Commons)

Heat pumps, air conditioners, and refrigerators utilize heat transfer from cold to hot. They are heat engines run backward. We say backward, rather than reverse, because except for Carnot engines, all heat engines, though they can be run backward, cannot truly be reversed. Heat transfer occurs from a cold reservoir Q_{c} and into a hot one. This requires work input W, which is also converted to heat transfer. Thus the heat transfer to the hot reservoir is $Q_{\mathrm{h}} = Q_{\mathrm{c}} + W$. (Note that Q_{h}, Q_{c}, and W are positive, with their directions indicated on schematics rather than by sign.) A heat pump's mission is for heat transfer Q_{h} to occur into a warm environment, such as a home in the winter. The mission of air conditioners and refrigerators is for heat transfer Q_c to occur from a cool environment, such as chilling a room or keeping food at lower temperatures than the environment. (Actually, a heat pump can be used both to heat and cool a space. It is essentially an air conditioner and a heating unit all in one. In this section we will concentrate on its heating mode.)

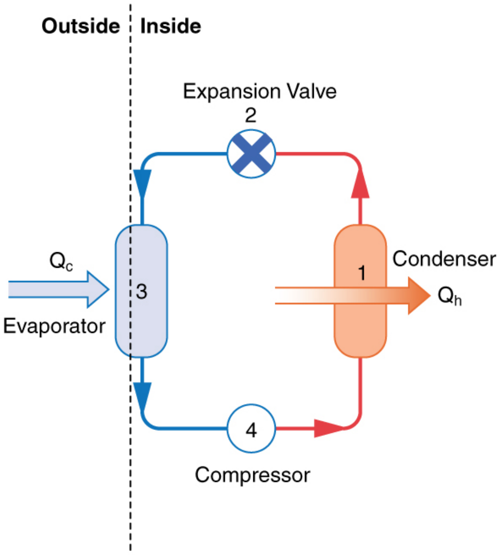


Heat pumps, air conditioners, and refrigerators are heat engines operated backward. The one shown here is based on a Carnot (reversible) engine. (a) Schematic diagram showing heat transfer from a cold reservoir to a warm reservoir with a heat pump. The directions of W, Q_h, and Q_c are opposite what they would be in a heat engine. (b) PV diagram for a Carnot cycle similar to that in [link] but reversed, following path ADCBA. The area inside the loop is negative, meaning there is a net work input. There is heat transfer Q_c into the system from a cold reservoir along path DC, and heat transfer Q_h out of the system into a hot reservoir along path BA.

Heat Pumps

The great advantage of using a heat pump to keep your home warm, rather than just burning fuel, is that a heat pump supplies $Q_\mathrm{h} = Q_\mathrm{c} + W$. Heat transfer is from the outside air, even at a temperature below freezing, to the indoor space. You only pay for W, and you get an additional heat transfer of Q_c from the outside at no cost; in many cases, at least twice as much energy is transferred to the heated space as is used to run the heat pump. When you burn fuel to keep warm, you pay for all of it. The disadvantage is that the work input (required by the second law of thermodynamics) is sometimes more expensive than simply burning fuel, especially if the work is done by electrical energy.

The basic components of a heat pump in its heating mode are shown in [link]. A working fluid such as a non-CFC refrigerant is used. In the outdoor coils (the evaporator), heat transfer Q_c occurs to the working fluid from the cold outdoor air, turning it into a gas.



A simple heat pump has four basic components: (1) condenser, (2) expansion valve, (3) evaporator, and (4) compressor. In the heating mode, heat transfer Q_c occurs to the working fluid in the evaporator (3) from the colder outdoor air, turning it into a gas. The electrically driven compressor (4) increases the temperature and pressure of the gas and forces it into the condenser coils (1) inside the heated space. Because the temperature of the gas is higher than the temperature in the room, heat transfer from the gas to the room occurs as the gas condenses to a liquid. The working fluid is then cooled as it flows back through an expansion valve (2) to the outdoor evaporator coils.

The electrically driven compressor (work input W) raises the temperature and pressure of the gas and forces it into the condenser coils that are inside the heated space. Because the temperature of the gas is higher than the temperature inside the room, heat transfer to the room occurs and the gas condenses to a liquid. The liquid then flows back through a pressure-reducing valve to the outdoor evaporator coils, being cooled through expansion. (In a cooling cycle, the evaporator and condenser coils exchange roles and the flow direction of the fluid is reversed.)

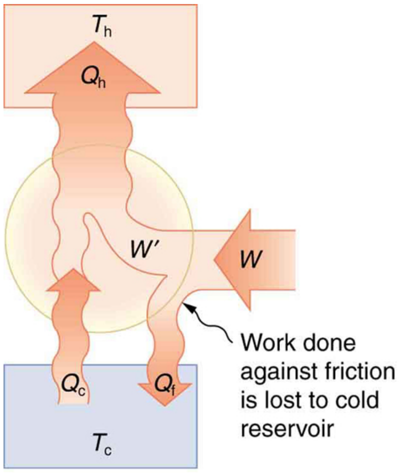
The quality of a heat pump is judged by how much heat transfer Q_h occurs into the warm space compared with how much work input W is required. In the spirit of taking the ratio of what you get to what you spend, we define a **heat pump's coefficient of performance** (COP_hp) to be

Equation:

$$COP_\mathrm{hp} = \frac{Q_\mathrm{h}}{W}.$$

Since the efficiency of a heat engine is $Eff = W/Q_\mathrm{h}$, we see that $COP_\mathrm{hp} = 1/Eff$, an important and interesting fact. First, since the efficiency of any heat engine is less than 1, it means that COP_hp is always greater than 1—that is, a heat pump always has more heat transfer Q_h than work put into it. Second, it means that heat pumps work best when temperature differences are small. The efficiency of a perfect, or Carnot, engine is $Eff_\mathrm{C} = 1 - (T_\mathrm{c}/T_\mathrm{h})$; thus, the smaller the temperature difference, the smaller the efficiency and the greater the COP_hp (because $COP_\mathrm{hp} = 1/Eff$). In other words, heat pumps do not work as well in very cold climates as they do in more moderate climates.

Friction and other irreversible processes reduce heat engine efficiency, but they do *not* benefit the operation of a heat pump—instead, they reduce the work input by converting part of it to heat transfer back into the cold reservoir before it gets into the heat pump.


T_h
Q_h
W'
W
Q_c
Q_f
T_c
Work done against friction is lost to cold reservoir

When a real heat engine is run backward, some of the intended work input (W) goes into heat transfer before it gets into the heat engine, thereby reducing its coefficient of performance COP_{hp}. In this figure, W’ represents the portion of W that goes into the heat pump, while the remainder of W is lost in the form of frictional heat (Q_f) to the cold reservoir. If all of W had gone into the heat pump, then Q_{h} would have been greater. The best heat pump uses adiabatic and isothermal processes, since, in theory, there would be no dissipative processes to reduce the heat transfer to the hot reservoir.

Example:

The Best *COP* $_{\text{hp}}$ of a Heat Pump for Home Use

A heat pump used to warm a home must employ a cycle that produces a working fluid at temperatures greater than typical indoor temperature so that heat transfer to the inside can take place. Similarly, it must produce a working fluid at temperatures that are colder than the outdoor temperature so that heat transfer occurs from outside. Its hot and cold reservoir temperatures therefore cannot be too close, placing a limit on its COP_{hp}. (See [link].) What is the best coefficient of performance possible for such a heat pump, if it has a hot reservoir temperature of 45.0^{o}C and a cold reservoir temperature of -15.0^{o}C?

Strategy

A Carnot engine reversed will give the best possible performance as a heat pump. As noted above, $COP_{\text{hp}} = 1/Eff$, so that we need to first calculate the Carnot efficiency to solve this problem.

Solution

Carnot efficiency in terms of absolute temperature is given by**:**

Equation:

$$Eff_{\text{C}} = 1 - \frac{T_{\text{c}}}{T_{\text{h}}}.$$

The temperatures in kelvins are $T_{\text{h}} = 318\ \text{K}$ and $T_{\text{c}} = 258\ \text{K}$, so that

Equation:

$$Eff_{\text{C}} = 1 - \frac{258\ \text{K}}{318\ \text{K}} = 0.1887.$$

Thus, from the discussion above,

Equation:

$$COP_{\text{hp}} = \frac{1}{\text{Eff}} = \frac{1}{0.1887} = 5.30,$$

or

Equation:

$$COP_{\text{hp}} = \frac{Q_{\text{h}}}{W} = 5.30,$$

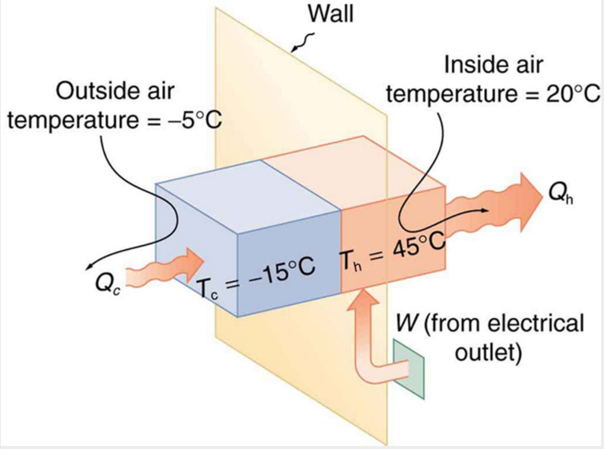
so that

Equation:

$$Q_{\text{h}} = 5.30\ \text{W}.$$

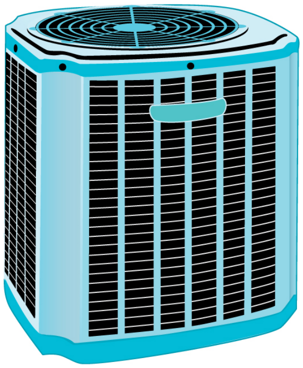
Discussion

This result means that the heat transfer by the heat pump is 5.30 times as much as the work put into it. It would cost 5.30 times as much for the same heat transfer by an electric room heater as it does for that produced by this heat pump. This is not a violation of conservation of energy. Cold ambient air provides 4.3 J per 1 J of work from the electrical outlet.


Wall
Outside air temperature = –5°C
Inside air temperature = 20°C
Q_h
Q_c
T_c = –15°C
T_h = 45°C
W (from electrical outlet)

Heat transfer from the outside to the inside, along with work done to run the pump, takes place in the heat pump of the example above. Note that the cold temperature produced by the heat pump is lower than the outside temperature, so that heat transfer into the working fluid occurs. The pump's compressor produces a temperature greater than the indoor temperature in order for heat transfer into the house to occur.

Real heat pumps do not perform quite as well as the ideal one in the previous example; their values of COP_{hp} range from about 2 to 4. This range means that the heat transfer Q_{h} from the heat pumps is 2 to 4 times as great as the work W put into them. Their economical feasibility is still limited, however, since W is usually supplied by electrical energy that costs more per joule than heat transfer by burning fuels like natural gas. Furthermore, the initial cost of a heat pump is greater than that of many furnaces, so that a heat pump must last longer for its cost to be recovered. Heat pumps are most likely to be economically superior where winter temperatures are mild, electricity is relatively cheap, and other fuels are relatively expensive. Also, since they can cool as well as heat a space, they have advantages where cooling in summer months is also desired. Thus some of the best locations for heat pumps are in warm summer climates with cool winters. [link] shows a heat pump, called a “*reverse cycle*” or “*split-system cooler*” in some countries.



In hot weather, heat transfer occurs from air inside the room to air outside, cooling the room. In cool weather, heat transfer occurs from air outside to air inside, warming the room. This switching is achieved by reversing the direction of flow of the working fluid.

Air Conditioners and Refrigerators

Air conditioners and refrigerators are designed to cool something down in a warm environment. As with heat pumps, work input is required for heat transfer from cold to hot, and this is expensive. The quality of air conditioners and refrigerators is judged by how much heat transfer Q_{c} occurs from a cold environment compared with how much work input W is required. What is considered the benefit in a heat pump is considered waste heat in a refrigerator. We thus define the **coefficient of performance** (COP_{ref}) of an air conditioner or refrigerator to be

Equation:

$$COP_{\text{ref}} = \frac{Q_{\text{c}}}{W}.$$

Noting again that $Q_{\text{h}} = Q_{\text{c}} + W$, we can see that an air conditioner will have a lower coefficient of performance than a heat pump, because $COP_{\text{hp}} = Q_{\text{h}}/W$ and Q_{h} is greater than Q_{c}. In this module's Problems and Exercises, you will show that

Equation:

$$COP_{\text{ref}} = COP_{\text{hp}} - 1$$

for a heat engine used as either an air conditioner or a heat pump operating between the same two temperatures. Real air conditioners and refrigerators typically do remarkably well, having values of COP_{ref} ranging from 2 to 6.

These numbers are better than the COP_{hp} values for the heat pumps mentioned above, because the temperature differences are smaller, but they are less than those for Carnot engines operating between the same two temperatures.

A type of COP rating system called the “energy efficiency rating” (EER) has been developed. This rating is an example where non-SI units are still used and relevant to consumers. To make it easier for the consumer, Australia, Canada, New Zealand, and the U.S. use an Energy Star Rating out of 5 stars—the more stars, the more energy efficient the appliance. EERs are expressed in mixed units of British thermal units (Btu) per hour of heating or cooling divided by the power input in watts. Room air conditioners are readily available with EERs ranging from 6 to 12. Although not the same as the COPs just described, these EERs are good for comparison purposes—the greater the EER, the cheaper an air conditioner is to operate (but the higher its purchase price is likely to be).

The EER of an air conditioner or refrigerator can be expressed as
Equation:

$$EER = \frac{Q_{\text{c}}/t_1}{W/t_2},$$

where Q_{c} is the amount of heat transfer from a cold environment in British thermal units, t_1 is time in hours, W is the work input in joules, and t_2 is time in seconds.

Note:
Problem-Solving Strategies for Thermodynamics

1. *Examine the situation to determine whether heat, work, or internal energy are involved.* Look for any system where the primary methods of transferring energy are heat and work. Heat engines, heat pumps, refrigerators, and air conditioners are examples of such systems.

2. *Identify the system of interest and draw a labeled diagram of the system showing energy flow.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns).* A written list is useful. Maximum efficiency means a Carnot engine is involved. Efficiency is not the same as the coefficient of performance.
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns).* Be sure to distinguish heat transfer into a system from heat transfer out of the system, as well as work input from work output. In many situations, it is useful to determine the type of process, such as isothermal or adiabatic.
5. *Solve the appropriate equation for the quantity to be determined (the unknown).*
6. *Substitute the known quantities along with their units into the appropriate equation and obtain numerical solutions complete with units.*
7. *Check the answer to see if it is reasonable: Does it make sense?* For example, efficiency is always less than 1, whereas coefficients of performance are greater than 1.

Section Summary

- An artifact of the second law of thermodynamics is the ability to heat an interior space using a heat pump. Heat pumps compress cold ambient air and, in so doing, heat it to room temperature without violation of conservation principles.
- To calculate the heat pump’s coefficient of performance, use the equation $COP_{\text{hp}} = \frac{Q_{\text{h}}}{W}$.
- A refrigerator is a heat pump; it takes warm ambient air and expands it to chill it.

Conceptual Questions

Exercise:

Problem:

Explain why heat pumps do not work as well in very cold climates as they do in milder ones. Is the same true of refrigerators?

Exercise:

Problem:

In some Northern European nations, homes are being built without heating systems of any type. They are very well insulated and are kept warm by the body heat of the residents. However, when the residents are not at home, it is still warm in these houses. What is a possible explanation?

Exercise:

Problem:

Why do refrigerators, air conditioners, and heat pumps operate most cost-effectively for cycles with a small difference between T_{h} and T_{c}? (Note that the temperatures of the cycle employed are crucial to its COP.)

Exercise:

Problem:

Grocery store managers contend that there is *less* total energy consumption in the summer if the store is kept at a *low* temperature. Make arguments to support or refute this claim, taking into account that there are numerous refrigerators and freezers in the store.

Exercise:

Problem:

Can you cool a kitchen by leaving the refrigerator door open?

Problem Exercises

Exercise:

Problem:

What is the coefficient of performance of an ideal heat pump that has heat transfer from a cold temperature of $-25.0^{\circ}\mathrm{C}$ to a hot temperature of $40.0^{\circ}\mathrm{C}$?

Solution:

4.82

Exercise:

Problem:

Suppose you have an ideal refrigerator that cools an environment at $-20.0^{\circ}\mathrm{C}$ and has heat transfer to another environment at $50.0^{\circ}\mathrm{C}$. What is its coefficient of performance?

Exercise:

Problem:

What is the best coefficient of performance possible for a hypothetical refrigerator that could make liquid nitrogen at $-200^{\circ}\mathrm{C}$ and has heat transfer to the environment at $35.0^{\circ}\mathrm{C}$?

Solution:

0.311

Exercise:

Problem:

In a very mild winter climate, a heat pump has heat transfer from an environment at $5.00^{\circ}\mathrm{C}$ to one at $35.0^{\circ}\mathrm{C}$. What is the best possible coefficient of performance for these temperatures? Explicitly show how you follow the steps in the Problem-Solving Strategies for Thermodynamics.

Exercise:

Problem:

(a) What is the best coefficient of performance for a heat pump that has a hot reservoir temperature of 50.0^{o}C and a cold reservoir temperature of -20.0^{o}C? (b) How much heat transfer occurs into the warm environment if $3.60 \times 10^7\ \text{J}$ of work ($10.0\text{kW} \cdot \text{h}$) is put into it? (c) If the cost of this work input is $10.0\ \text{cents/kW} \cdot \text{h}$, how does its cost compare with the direct heat transfer achieved by burning natural gas at a cost of 85.0 cents per therm. (A therm is a common unit of energy for natural gas and equals $1.055 \times 10^8\ \text{J}$.)

Solution:

(a) 4.61

(b) $1.66 \times 10^8\ \text{J}$ or $3.97 \times 10^4\ \text{kcal}$

(c) To transfer $1.66 \times 10^8\ \text{J}$, heat pump costs $1.00, natural gas costs $1.34.

Exercise:

Problem:

(a) What is the best coefficient of performance for a refrigerator that cools an environment at -30.0^{o}C and has heat transfer to another environment at 45.0^{o}C? (b) How much work in joules must be done for a heat transfer of 4186 kJ from the cold environment? (c) What is the cost of doing this if the work costs 10.0 cents per $3.60 \times 10^6\ \text{J}$ (a kilowatt-hour)? (d) How many kJ of heat transfer occurs into the warm environment? (e) Discuss what type of refrigerator might operate between these temperatures.

Exercise:

Problem:

Suppose you want to operate an ideal refrigerator with a cold temperature of -10.0°C, and you would like it to have a coefficient of performance of 7.00. What is the hot reservoir temperature for such a refrigerator?

Solution:

27.6°C

Exercise:

Problem:

An ideal heat pump is being considered for use in heating an environment with a temperature of 22.0°C. What is the cold reservoir temperature if the pump is to have a coefficient of performance of 12.0?

Exercise:

Problem:

A 4-ton air conditioner removes $5.06\times10^{7}\text{ J}$ (48,000 British thermal units) from a cold environment in 1.00 h. (a) What energy input in joules is necessary to do this if the air conditioner has an energy efficiency rating (EER) of 12.0? (b) What is the cost of doing this if the work costs 10.0 cents per $3.60\times10^{6}\text{ J}$ (one kilowatt-hour)? (c) Discuss whether this cost seems realistic. Note that the energy efficiency rating (EER) of an air conditioner or refrigerator is defined to be the number of British thermal units of heat transfer from a cold environment per hour divided by the watts of power input.

Solution:

(a) $1.44\times10^{7}\text{ J}$

(b) 40 cents

(c) This cost seems quite realistic; it says that running an air conditioner all day would cost $9.59 (if it ran continuously).

Exercise:

Problem:

Show that the coefficients of performance of refrigerators and heat pumps are related by $COP_{\text{ref}} = COP_{\text{hp}} - 1$.

Start with the definitions of the COP s and the conservation of energy relationship between Q_{h}, Q_{c}, and W.

Glossary

heat pump
: a machine that generates heat transfer from cold to hot

coefficient of performance
: for a heat pump, it is the ratio of heat transfer at the output (the hot reservoir) to the work supplied; for a refrigerator or air conditioner, it is the ratio of heat transfer from the cold reservoir to the work supplied

Entropy and the Second Law of Thermodynamics: Disorder and the Unavailability of Energy

- Define entropy and calculate the increase of entropy in a system with reversible and irreversible processes.
- Explain the expected fate of the universe in entropic terms.
- Calculate the increasing disorder of a system.

The ice in this drink is slowly melting. Eventually the liquid will reach thermal equilibrium, as predicted by the second law of thermodynamics. (credit: Jon Sullivan, PDPhoto.org)

There is yet another way of expressing the second law of thermodynamics. This version relates to a concept called **entropy**. By examining it, we shall see that the directions associated with the second law—heat transfer from hot to cold, for example—are related to the tendency in nature for systems to become disordered and for less energy to be available for use as work. The entropy of a system can in fact be shown to be a measure of its disorder and of the unavailability of energy to do work.

Note:

Making Connections: Entropy, Energy, and Work

Recall that the simple definition of energy is the ability to do work. Entropy is a measure of how much energy is not available to do work. Although all forms of energy are interconvertible, and all can be used to do work, it is not always possible, even in principle, to convert the entire available energy into work. That unavailable energy is of interest in thermodynamics, because the field of thermodynamics arose from efforts to convert heat to work.

We can see how entropy is defined by recalling our discussion of the Carnot engine. We noted that for a Carnot cycle, and hence for any reversible processes, $Q_{\text{c}}/Q_{\text{h}} = T_{\text{c}}/T_{\text{h}}$. Rearranging terms yields

Equation:

$$\frac{Q_{\text{c}}}{T_{\text{c}}} = \frac{Q_{\text{h}}}{T_{\text{h}}}$$

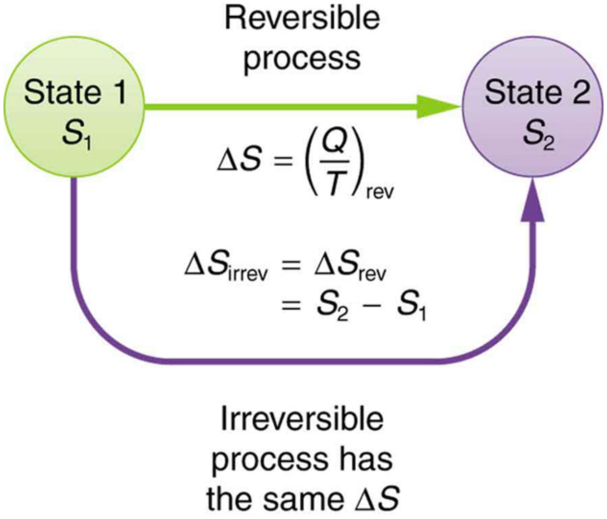
for any reversible process. Q_{c} and Q_{h} are absolute values of the heat transfer at temperatures T_{c} and T_{h}, respectively. This ratio of Q/T is defined to be the **change in entropy** ΔS for a reversible process,

Equation:

$$\Delta S = \left(\frac{Q}{T}\right)_{\text{rev}},$$

where Q is the heat transfer, which is positive for heat transfer into and negative for heat transfer out of, and T is the absolute temperature at which the reversible process takes place. The SI unit for entropy is joules per kelvin (J/K). If temperature changes during the process, then it is usually a good approximation (for small changes in temperature) to take T to be the average temperature, avoiding the need to use integral calculus to find ΔS.

The definition of ΔS is strictly valid only for reversible processes, such as used in a Carnot engine. However, we can find ΔS precisely even for real, irreversible processes. The reason is that the entropy S of a system, like internal energy U, depends only on the state of the system and not how it reached that condition. Entropy is a property of state. Thus the change in entropy ΔS of a system between state 1 and state 2 is the same no matter how the change occurs. We just need to find or imagine a reversible process that takes us from state 1 to state 2 and calculate ΔS for that process. That will be the change in entropy for any process going from state 1 to state 2. (See [link].)


Reversible process
State 1
S_1
State 2
S_2
$\Delta S = \left(\frac{Q}{T}\right)_{\text{rev}}$
$\Delta S_{\text{irrev}} = \Delta S_{\text{rev}}$
$= S_2 - S_1$
Irreversible process has the same ΔS

When a system goes from state 1 to state 2, its entropy changes by the same amount ΔS, whether a hypothetical reversible path is followed or a real irreversible path is taken.

Now let us take a look at the change in entropy of a Carnot engine and its heat reservoirs for one full cycle. The hot reservoir has a loss of entropy $\Delta S_{\mathrm{h}} = -Q_{\mathrm{h}}/T_{\mathrm{h}}$, because heat transfer occurs out of it (remember that when heat transfers out, then Q has a negative sign). The cold reservoir has a gain of entropy $\Delta S_{\mathrm{c}} = Q_{\mathrm{c}}/T_{\mathrm{c}}$, because heat transfer occurs into it. (We assume the reservoirs are sufficiently large that their temperatures are constant.) So the total change in entropy is

Equation:

$$\Delta S_{\mathrm{tot}} = \Delta S_{\mathrm{h}} + \Delta S_{\mathrm{c}}.$$

Thus, since we know that $Q_{\mathrm{h}}/T_{\mathrm{h}} = Q_{\mathrm{c}}/T_{\mathrm{c}}$ for a Carnot engine,

Equation:

$$\Delta S_{\mathrm{tot}}{=}{-}\frac{Q_{\mathrm{h}}}{T_{\mathrm{h}}} + \frac{Q_{\mathrm{c}}}{T_{\mathrm{c}}} = 0.$$

This result, which has general validity, means that *the total change in entropy for a system in any reversible process is zero.*

The entropy of various parts of the system may change, but the total change is zero. Furthermore, the system does not affect the entropy of its surroundings, since heat transfer between them does not occur. Thus the reversible process changes neither the total entropy of the system nor the entropy of its surroundings. Sometimes this is stated as follows: *Reversible processes do not affect the total entropy of the universe.* Real processes are not reversible, though, and they do change total entropy. We can, however, use hypothetical reversible processes to determine the value of entropy in real, irreversible processes. The following example illustrates this point.

Example:

Entropy Increases in an Irreversible (Real) Process

Spontaneous heat transfer from hot to cold is an irreversible process. Calculate the total change in entropy if 4000 J of heat transfer occurs from a hot reservoir at $T_{\text{h}} = 600\text{ K}(327^{\text{o}}\text{ C})$ to a cold reservoir at $T_{\text{c}} = 250\text{ K}(-23^{\text{o}}\text{ C})$, assuming there is no temperature change in either reservoir. (See [link].)

Strategy

How can we calculate the change in entropy for an irreversible process when $\Delta S_{\text{tot}} = \Delta S_{\text{h}} + \Delta S_{\text{c}}$ is valid only for reversible processes? Remember that the total change in entropy of the hot and cold reservoirs will be the same whether a reversible or irreversible process is involved in heat transfer from hot to cold. So we can calculate the change in entropy of the hot reservoir for a hypothetical reversible process in which 4000 J of heat transfer occurs from it; then we do the same for a hypothetical reversible process in which 4000 J of heat transfer occurs to the cold reservoir. This produces the same changes in the hot and cold reservoirs that would occur if the heat transfer were allowed to occur irreversibly between them, and so it also produces the same changes in entropy.

Solution

We now calculate the two changes in entropy using $\Delta S_{\text{tot}} = \Delta S_{\text{h}} + \Delta S_{\text{c}}$. First, for the heat transfer from the hot reservoir,

Equation:

$$\Delta S_{\text{h}} = \frac{-Q_{\text{h}}}{T_{\text{h}}} = \frac{-4000\text{ J}}{600\text{ K}} = -6.67\text{ J/K}.$$

And for the cold reservoir,

Equation:

$$\Delta S_{\text{c}} = \frac{Q_{\text{c}}}{T_{\text{c}}} = \frac{4000\text{ J}}{250\text{ K}} = 16.0\text{ J/K}.$$

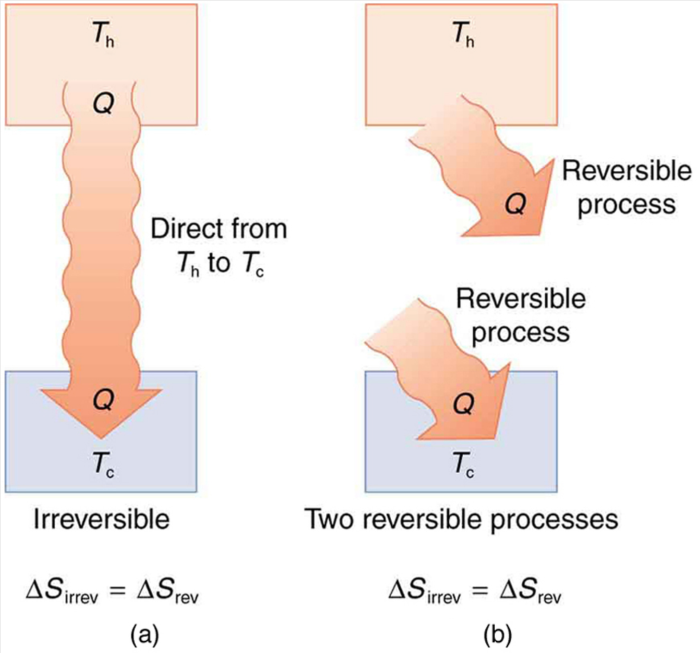
Thus the total is

Equation:

$$\begin{array}{rcl} \Delta S_{\text{tot}} & = & \Delta S_{\text{h}} + \Delta S_{\text{c}} \\ & = & (-6.67 + 16.0)\text{ J/K} \\ & = & 9.33\text{ J/K}. \end{array}$$

Discussion

There is an *increase* in entropy for the system of two heat reservoirs undergoing this irreversible heat transfer. We will see that this means there is a loss of ability to do work with this transferred energy. Entropy has increased, and energy has become unavailable to do work.


T_h
Q
Direct from T_h to T_c
Q
T_c
Irreversible
$\Delta S_{irrev} = \Delta S_{rev}$
(a)
T_h
Reversible process
Q
Reversible process
Q
T_c
Two reversible processes
$\Delta S_{irrev} = \Delta S_{rev}$
(b)

(a) Heat transfer from a hot object to a cold one is an irreversible process that produces an overall increase in entropy. (b) The same final state and, thus, the same change in entropy is achieved for the objects if reversible heat transfer processes occur between the two objects whose temperatures are the same as the temperatures of the corresponding objects in the irreversible process.

It is reasonable that entropy increases for heat transfer from hot to cold. Since the change in entropy is Q/T, there is a larger change at lower temperatures. The decrease in entropy of the hot object is therefore less than the increase in entropy of the cold object, producing an overall increase, just as in the previous example. This result is very general:

There is an increase in entropy for any system undergoing an irreversible process.

With respect to entropy, there are only two possibilities: entropy is constant for a reversible process, and it increases for an irreversible process. There is a fourth version of **the second law of thermodynamics stated in terms of entropy**:

The total entropy of a system either increases or remains constant in any process; it never decreases.

For example, heat transfer cannot occur spontaneously from cold to hot, because entropy would decrease.

Entropy is very different from energy. Entropy is *not* conserved but increases in all real processes. Reversible processes (such as in Carnot engines) are the processes in which the most heat transfer to work takes place and are also the ones that keep entropy constant. Thus we are led to make a connection between entropy and the availability of energy to do work.

Entropy and the Unavailability of Energy to Do Work

What does a change in entropy mean, and why should we be interested in it? One reason is that entropy is directly related to the fact that not all heat transfer can be converted into work. The next example gives some indication of how an increase in entropy results in less heat transfer into work.

Example:
Less Work is Produced by a Given Heat Transfer When Entropy Change is Greater

(a) Calculate the work output of a Carnot engine operating between temperatures of 600 K and 100 K for 4000 J of heat transfer to the engine. (b) Now suppose that the 4000 J of heat transfer occurs first from the 600 K reservoir to a 250 K reservoir (without doing any work, and this produces the increase in entropy calculated above) before transferring into a Carnot engine operating between 250 K and 100 K. What work output is produced? (See [link].)

Strategy

In both parts, we must first calculate the Carnot efficiency and then the work output.

Solution (a)

The Carnot efficiency is given by

Equation:

$$Eff_{\text{C}} = 1 - \frac{T_{\text{c}}}{T_{\text{h}}}.$$

Substituting the given temperatures yields

Equation:

$$Eff_{\text{C}} = 1 - \frac{100\text{ K}}{600\text{ K}} = 0.833.$$

Now the work output can be calculated using the definition of efficiency for any heat engine as given by

Equation:

$$\text{Eff} = \frac{W}{Q_{\text{h}}}.$$

Solving for W and substituting known terms gives

Equation:

$$\begin{array}{rcl} W & = & Eff_{\text{C}}Q_{\text{h}} \\ & = & (0.833)(4000\text{ J}) = 3333\text{ J}. \end{array}$$

Solution (b)

Similarly,

Equation:

$$Eff\prime_{\mathrm{C}} = 1 - \frac{T_{\mathrm{c}}}{T\prime_{\mathrm{c}}} = 1 - \frac{100\ \mathrm{K}}{250\ \mathrm{K}} = 0.600,$$

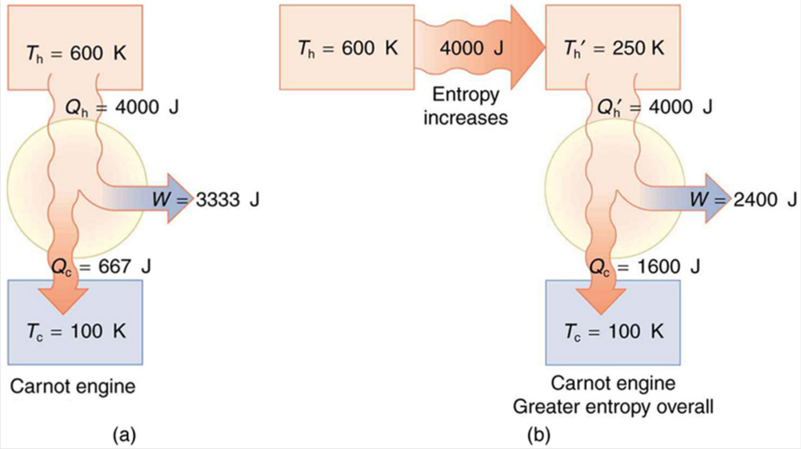
so that

Equation:

$$\begin{aligned} W &= Eff\prime_{\mathrm{C}} Q_h \\ &= (0.600)(4000\ \mathrm{J}) = 2400\ \mathrm{J}. \end{aligned}$$

Discussion

There is 933 J less work from the same heat transfer in the second process. This result is important. The same heat transfer into two perfect engines produces different work outputs, because the entropy change differs in the two cases. In the second case, entropy is greater and less work is produced. Entropy is associated with the *un*availability of energy to do work.


T_h = 600 K
Q_h = 4000 J
W = 3333 J
Q_c = 667 J
T_c = 100 K
Carnot engine
(a)
T_h = 600 K
4000 J
Entropy increases
T_h' = 250 K
Q_h' = 4000 J
W = 2400 J
Q_c = 1600 J
T_c = 100 K
Carnot engine
Greater entropy overall
(b)

(a) A Carnot engine working at between 600 K and 100 K has 4000 J of heat transfer and performs 3333 J of work. (b) The 4000 J of heat transfer occurs first irreversibly to a 250 K reservoir and then goes into a Carnot engine. The increase in entropy caused by the heat transfer to a colder reservoir results in a smaller work output of 2400 J. There is a permanent loss of 933 J of energy for the purpose of doing work.

When entropy increases, a certain amount of energy becomes *permanently* unavailable to do work. The energy is not lost, but its character is changed, so that some of it can never be converted to doing work—that is, to an organized force acting through a distance. For instance, in the previous example, 933 J less work was done after an increase in entropy of 9.33 J/K occurred in the 4000 J heat transfer from the 600 K reservoir to the 250 K reservoir. It can be shown that the amount of energy that becomes unavailable for work is

Equation:

$$W_{\text{unavail}} = \Delta S \cdot T_0,$$

where T_0 is the lowest temperature utilized. In the previous example,

Equation:

$$W_{\text{unavail}} = (9.33 \text{ J/K})(100 \text{ K}) = 933 \text{ J}$$

as found.

Heat Death of the Universe: An Overdose of Entropy

In the early, energetic universe, all matter and energy were easily interchangeable and identical in nature. Gravity played a vital role in the young universe. Although it may have *seemed* disorderly, and therefore, superficially entropic, in fact, there was enormous potential energy available to do work—all the future energy in the universe.

As the universe matured, temperature differences arose, which created more opportunity for work. Stars are hotter than planets, for example, which are warmer than icy asteroids, which are warmer still than the vacuum of the space between them.

Most of these are cooling down from their usually violent births, at which time they were provided with energy of their own—nuclear energy in the case of stars, volcanic energy on Earth and other planets, and so on. Without additional energy input, however, their days are numbered.

As entropy increases, less and less energy in the universe is available to do work. On Earth, we still have great stores of energy such as fossil and nuclear fuels; large-scale temperature differences, which can provide wind energy; geothermal energies due to differences in temperature in Earth’s layers; and tidal energies owing to our abundance of liquid water. As these are used, a certain fraction of the energy they contain can never be converted into doing work. Eventually, all fuels will be exhausted, all temperatures will equalize, and it will be impossible for heat engines to function, or for work to be done.

Entropy increases in a closed system, such as the universe. But in parts of the universe, for instance, in the Solar system, it is not a locally closed system. Energy flows from the Sun to the planets, replenishing Earth’s stores of energy. The Sun will continue to supply us with energy for about another five billion years. We will enjoy direct solar energy, as well as side effects of solar energy, such as wind power and biomass energy from photosynthetic plants. The energy from the Sun will keep our water at the liquid state, and the Moon’s gravitational pull will continue to provide tidal energy. But Earth’s geothermal energy will slowly run down and won’t be replenished.

But in terms of the universe, and the very long-term, very large-scale picture, the entropy of the universe is increasing, and so the availability of energy to do work is constantly decreasing. Eventually, when all stars have died, all forms of potential energy have been utilized, and all temperatures have equalized (depending on the mass of the universe, either at a very high temperature following a universal contraction, or a very low one, just before all activity ceases) there will be no possibility of doing work.

Either way, the universe is destined for thermodynamic equilibrium—maximum entropy. This is often called the *heat death of the universe*, and will mean the end of all activity. However, whether the universe contracts and heats up, or continues to expand and cools down, the end is not near.

Calculations of black holes suggest that entropy can easily continue for at least 10^{100} years.

Order to Disorder

Entropy is related not only to the unavailability of energy to do work—it is also a measure of disorder. This notion was initially postulated by Ludwig Boltzmann in the 1800s. For example, melting a block of ice means taking a highly structured and orderly system of water molecules and converting it into a disorderly liquid in which molecules have no fixed positions. (See [link].) There is a large increase in entropy in the process, as seen in the following example.

Example:

Entropy Associated with Disorder

Find the increase in entropy of 1.00 kg of ice originally at $0^{\text{o}}\ \text{C}$ that is melted to form water at $0^{\text{o}}\ \text{C}$.

Strategy

As before, the change in entropy can be calculated from the definition of ΔS once we find the energy Q needed to melt the ice.

Solution

The change in entropy is defined as:

Equation:

$$\mathit{\Delta}S=\frac{Q}{T}.$$

Here Q is the heat transfer necessary to melt 1.00 kg of ice and is given by

Equation:

$$Q=mL_{\text{f}},$$

where m is the mass and L_{f} is the latent heat of fusion. $L_{\text{f}}=334\ \text{kJ/kg}$ for water, so that

Equation:

$$Q = (1.00\text{ kg})(334\text{ kJ/kg}) = 3.34 \times 10^5\text{ J}.$$

Now the change in entropy is positive, since heat transfer occurs into the ice to cause the phase change; thus,

Equation:

$$\Delta S = \frac{Q}{T} = \frac{3.34 \times 10^5\text{ J}}{T}.$$

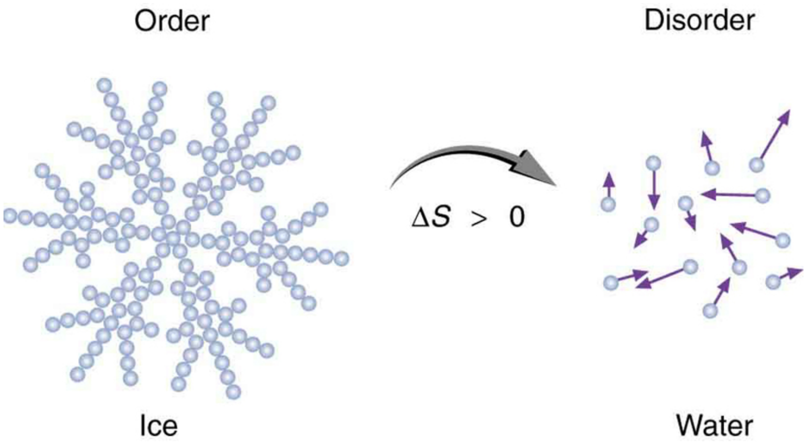
T is the melting temperature of ice. That is, $T = 0^{\text{o}}\text{C}{=}273\text{ K}$. So the change in entropy is

Equation:

$$\begin{array}{rcl}\Delta S & = & \frac{3.34{\times}10^5\text{ J}}{273\text{ K}} \\ & = & 1.22{\times}10^3\text{ J/K}.\end{array}$$

Discussion

This is a significant increase in entropy accompanying an increase in disorder.



When ice melts, it becomes more disordered and less structured. The systematic arrangement of molecules in a crystal structure is replaced by a more random and less orderly movement of molecules without fixed locations or orientations. Its entropy increases because heat transfer occurs into it. Entropy is a measure of disorder.

In another easily imagined example, suppose we mix equal masses of water originally at two different temperatures, say 20.0° C and 40.0° C. The result is water at an intermediate temperature of 30.0° C. Three outcomes have resulted: entropy has increased, some energy has become unavailable to do work, and the system has become less orderly. Let us think about each of these results.

First, entropy has increased for the same reason that it did in the example above. Mixing the two bodies of water has the same effect as heat transfer from the hot one and the same heat transfer into the cold one. The mixing decreases the entropy of the hot water but increases the entropy of the cold water by a greater amount, producing an overall increase in entropy.

Second, once the two masses of water are mixed, there is only one temperature—you cannot run a heat engine with them. The energy that could have been used to run a heat engine is now unavailable to do work.

Third, the mixture is less orderly, or to use another term, less structured. Rather than having two masses at different temperatures and with different distributions of molecular speeds, we now have a single mass with a uniform temperature.

These three results—entropy, unavailability of energy, and disorder—are not only related but are in fact essentially equivalent.

Life, Evolution, and the Second Law of Thermodynamics

Some people misunderstand the second law of thermodynamics, stated in terms of entropy, to say that the process of the evolution of life violates this law. Over time, complex organisms evolved from much simpler ancestors, representing a large decrease in entropy of the Earth’s biosphere. It is a fact that living organisms have evolved to be highly structured, and much lower in entropy than the substances from which they grow. But it is *always* possible for the entropy of one part of the universe to decrease, provided the total change in entropy of the universe increases. In equation form, we can write this as

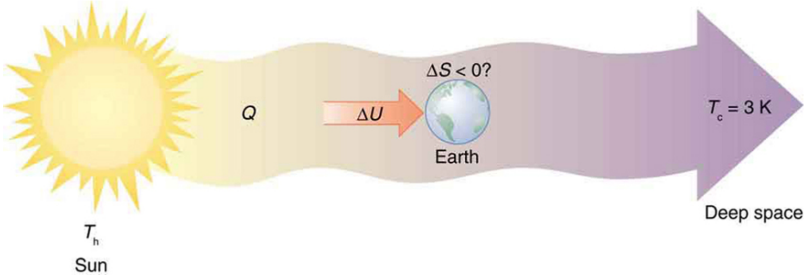
Equation:

$$\Delta S_{\text{tot}} = \Delta S_{\text{syst}} + \Delta S_{\text{envir}} > 0.$$

Thus ΔS_{syst} can be negative as long as ΔS_{envir} is positive and greater in magnitude.

How is it possible for a system to decrease its entropy? Energy transfer is necessary. If I pick up marbles that are scattered about the room and put them into a cup, my work has decreased the entropy of that system. If I gather iron ore from the ground and convert it into steel and build a bridge, my work has decreased the entropy of that system. Energy coming from the Sun can decrease the entropy of local systems on Earth—that is, ΔS_{syst} is negative. But the overall entropy of the rest of the universe increases by a greater amount—that is, ΔS_{envir} is positive and greater in magnitude. Thus, $\Delta S_{\text{tot}} = \Delta S_{\text{syst}} + \Delta S_{\text{envir}} > 0$, and the second law of thermodynamics is *not* violated.

Every time a plant stores some solar energy in the form of chemical potential energy, or an updraft of warm air lifts a soaring bird, the Earth can be viewed as a heat engine operating between a hot reservoir supplied by the Sun and a cold reservoir supplied by dark outer space—a heat engine of high complexity, causing local decreases in entropy as it uses part of the heat transfer from the Sun into deep space. There is a large total increase in entropy resulting from this massive heat transfer. A small part of this heat transfer is stored in structured systems on Earth, producing much smaller local decreases in entropy. (See [link].)



Earth's entropy may decrease in the process of intercepting a small part of the heat transfer from the Sun into deep space. Entropy for the entire process increases greatly while Earth becomes more structured with living systems and stored energy in various forms.

Note:

PhET Explorations: Reversible Reactions

Watch a reaction proceed over time. How does total energy affect a reaction rate? Vary temperature, barrier height, and potential energies. Record concentrations and time in order to extract rate coefficients. Do temperature dependent studies to extract Arrhenius parameters. This simulation is best used with teacher guidance because it presents an analogy of chemical reactions.

Reversible Reactions

Section Summary

- Entropy is the loss of energy available to do work.
- Another form of the second law of thermodynamics states that the total entropy of a system either increases or remains constant; it never decreases.
- Entropy is zero in a reversible process; it increases in an irreversible process.
- The ultimate fate of the universe is likely to be thermodynamic equilibrium, where the universal temperature is constant and no energy is available to do work.
- Entropy is also associated with the tendency toward disorder in a closed system.

Conceptual Questions

Exercise:

Problem:

A woman shuts her summer cottage up in September and returns in June. No one has entered the cottage in the meantime. Explain what she is likely to find, in terms of the second law of thermodynamics.

Exercise:

Problem:

Consider a system with a certain energy content, from which we wish to extract as much work as possible. Should the system's entropy be high or low? Is this orderly or disorderly? Structured or uniform? Explain briefly.

Exercise:

Problem:

Does a gas become more orderly when it liquefies? Does its entropy change? If so, does the entropy increase or decrease? Explain your answer.

Exercise:

Problem:

Explain how water's entropy can decrease when it freezes without violating the second law of thermodynamics. Specifically, explain what happens to the entropy of its surroundings.

Exercise:

Problem:

Is a uniform-temperature gas more or less orderly than one with several different temperatures? Which is more structured? In which can heat transfer result in work done without heat transfer from another system?

Exercise:

Problem:

Give an example of a spontaneous process in which a system becomes less ordered and energy becomes less available to do work. What happens to the system's entropy in this process?

Exercise:

Problem:

What is the change in entropy in an adiabatic process? Does this imply that adiabatic processes are reversible? Can a process be precisely adiabatic for a macroscopic system?

Exercise:

Problem:

Does the entropy of a star increase or decrease as it radiates? Does the entropy of the space into which it radiates (which has a temperature of about 3 K) increase or decrease? What does this do to the entropy of the universe?

Exercise:

Problem:

Explain why a building made of bricks has smaller entropy than the same bricks in a disorganized pile. Do this by considering the number of ways that each could be formed (the number of microstates in each macrostate).

Problem Exercises

Exercise:

Problem:

(a) On a winter day, a certain house loses $5.00\times10^{8}\text{ J}$ of heat to the outside (about 500,000 Btu). What is the total change in entropy due to this heat transfer alone, assuming an average indoor temperature of 21.0^{o} C and an average outdoor temperature of 5.00^{o} C? (b) This large change in entropy implies a large amount of energy has become unavailable to do work. Where do we find more energy when such energy is lost to us?

Solution:

(a) $9.78\times10^{4}\text{ J/K}$

(b) In order to gain more energy, we must generate it from things within the house, like a heat pump, human bodies, and other appliances. As you know, we use a lot of energy to keep our houses warm in the winter because of the loss of heat to the outside.

Exercise:

Problem:

On a hot summer day, $4.00\times10^{6}\text{ J}$ of heat transfer into a parked car takes place, increasing its temperature from 35.0^{o} C to 45.0^{o} C. What is the increase in entropy of the car due to this heat transfer alone?

Exercise:

Problem:

A hot rock ejected from a volcano’s lava fountain cools from 1100^{o} C to 40.0^{o} C, and its entropy decreases by 950 J/K. How much heat transfer occurs from the rock?

Solution:

$8.01 \times 10^5 \text{ J}$

Exercise:

Problem:

When $1.60{\times}10^5\text{ J}$ of heat transfer occurs into a meat pie initially at 20.0^{o} C, its entropy increases by 480 J/K. What is its final temperature?

Exercise:

Problem:

The Sun radiates energy at the rate of $3.80{\times}10^{26}\text{ W}$ from its 5500^{o} C surface into dark empty space (a negligible fraction radiates onto Earth and the other planets). The effective temperature of deep space is -270^{o} C. (a) What is the increase in entropy in one day due to this heat transfer? (b) How much work is made unavailable?

Solution:

(a) $1.04 \times 10^{31}\text{ J/K}$

(b) $3.28 \times 10^{31}\text{ J}$

Exercise:

Problem:

(a) In reaching equilibrium, how much heat transfer occurs from 1.00 kg of water at 40.0° C when it is placed in contact with 1.00 kg of 20.0° C water in reaching equilibrium? (b) What is the change in entropy due to this heat transfer? (c) How much work is made unavailable, taking the lowest temperature to be 20.0° C? Explicitly show how you follow the steps in the Problem-Solving Strategies for Entropy.

Exercise:

Problem:

What is the decrease in entropy of 25.0 g of water that condenses on a bathroom mirror at a temperature of 35.0° C, assuming no change in temperature and given the latent heat of vaporization to be 2450 kJ/kg?

Solution:

199 J/K

Exercise:

Problem:

Find the increase in entropy of 1.00 kg of liquid nitrogen that starts at its boiling temperature, boils, and warms to 20.0° C at constant pressure.

Exercise:

Problem:

A large electrical power station generates 1000 MW of electricity with an efficiency of 35.0%. (a) Calculate the heat transfer to the power station, Q_{h}, in one day. (b) How much heat transfer Q_{c} occurs to the environment in one day? (c) If the heat transfer in the cooling towers is from 35.0^{o} C water into the local air mass, which increases in temperature from 18.0^{o} C to 20.0^{o} C, what is the total increase in entropy due to this heat transfer? (d) How much energy becomes unavailable to do work because of this increase in entropy, assuming an 18.0^{o} C lowest temperature? (Part of Q_{c} could be utilized to operate heat engines or for simply heating the surroundings, but it rarely is.)

Solution:

(a) $2.47 \times 10^{14}\text{ J}$

(b) $1.60 \times 10^{14}\text{ J}$

(c) $2.85 \times 10^{10}\text{ J/K}$

(d) $8.29 \times 10^{12}\text{ J}$

Exercise:

Problem:

(a) How much heat transfer occurs from 20.0 kg of 90.0° C water placed in contact with 20.0 kg of 10.0° C water, producing a final temperature of 50.0° C? (b) How much work could a Carnot engine do with this heat transfer, assuming it operates between two reservoirs at constant temperatures of 90.0° C and 10.0° C? (c) What increase in entropy is produced by mixing 20.0 kg of 90.0° C water with 20.0 kg of 10.0° C water? (d) Calculate the amount of work made unavailable by this mixing using a low temperature of 10.0° C, and compare it with the work done by the Carnot engine. Explicitly show how you follow the steps in the Problem-Solving Strategies for Entropy. (e) Discuss how everyday processes make increasingly more energy unavailable to do work, as implied by this problem.

Glossary

entropy
: a measurement of a system's disorder and its inability to do work in a system

change in entropy
: the ratio of heat transfer to temperature Q/T

second law of thermodynamics stated in terms of entropy
: the total entropy of a system either increases or remains constant; it never decreases

Statistical Interpretation of Entropy and the Second Law of Thermodynamics: The Underlying Explanation

- Identify probabilities in entropy.
- Analyze statistical probabilities in entropic systems.

When you toss a coin a large number of times, heads and tails tend to come up in roughly equal numbers. Why doesn’t heads come up 100, 90, or even 80% of the time? (credit: Jon Sullivan, PDPhoto.org)

The various ways of formulating the second law of thermodynamics tell what happens rather than why it happens. Why should heat transfer occur only from hot to cold? Why should energy become ever less available to do work? Why should the universe become increasingly disorderly? The answer is that it is a matter of overwhelming probability. Disorder is simply vastly more likely than order.

When you watch an emerging rain storm begin to wet the ground, you will notice that the drops fall in a disorganized manner both in time and in space. Some fall close together, some far apart, but they never fall in straight, orderly rows. It is not impossible for rain to fall in an orderly pattern, just highly unlikely, because there are many more disorderly ways than orderly ones. To illustrate this fact, we will examine some random processes, starting with coin tosses.

Coin Tosses

What are the possible outcomes of tossing 5 coins? Each coin can land either heads or tails. On the large scale, we are concerned only with the total heads and tails and not with the order in which heads and tails appear. The following possibilities exist:

Equation:

$$\begin{gathered}\text{5 heads, 0 tails}\\ \text{4 heads, 1 tail}\\ \text{3 heads, 2 tails}\\ \text{2 heads, 3 tails}\\ \text{1 head, 4 tails}\\ \text{0 head, 5 tails}\end{gathered}$$

These are what we call macrostates. A **macrostate** is an overall property of a system. It does not specify the details of the system, such as the order in which heads and tails occur or which coins are heads or tails.

Using this nomenclature, a system of 5 coins has the 6 possible macrostates just listed. Some macrostates are more likely to occur than others. For instance, there is only one way to get 5 heads, but there are several ways to get 3 heads and 2 tails, making the latter macrostate more probable. [link] lists of all the ways in which 5 coins can be tossed, taking into account the order in which heads and tails occur. Each sequence is called a **microstate** —a detailed description of every element of a system.

	Individual microstates	**Number of microstates**
5 heads, 0 tails	HHHHH	1
4 heads, 1 tail	HHHHT, HHHTH, HHTHH, HTHHH, THHHH	5
3 heads, 2 tails	HTHTH, THTHH, HTHHT, THHTH, THHHT HTHTH, THTHH, HTHHT, THHTH, THHHT	10
2 heads, 3 tails	TTTHH, TTHHT, THHTT, HHTTT, TTHTH, THTHT, HTHTT, THTTH, HTTHT, HTTTH	10
1 head, 4 tails	TTTTH, TTTHT, TTHTT, THTTT, HTTTT	5
0 heads, 5 tails	TTTTT	1
		Total: 32

5-Coin Toss

The macrostate of 3 heads and 2 tails can be achieved in 10 ways and is thus 10 times more probable than the one having 5 heads. Not surprisingly, it is equally probable to have the reverse, 2 heads and 3 tails. Similarly, it is equally probable to get 5 tails as it is to get 5 heads. Note that all of these conclusions are based on the crucial assumption that each microstate is equally probable. With coin tosses, this requires that the coins not be asymmetric in a way that favors one side over the other, as with loaded dice. With any system, the assumption that all microstates are equally probable must be valid, or the analysis will be erroneous.

The two most orderly possibilities are 5 heads or 5 tails. (They are more structured than the others.) They are also the least likely, only 2 out of 32 possibilities. The most disorderly possibilities are 3 heads and 2 tails and its reverse. (They are the least structured.) The most disorderly possibilities are also the most likely, with 20 out of 32 possibilities for the 3 heads and 2 tails and its reverse. If we start with an orderly array like 5 heads and toss the coins, it is very likely that we will get a less orderly array as a result, since 30 out of the 32 possibilities are less orderly. So even if you start with an orderly state, there is a strong tendency to go from order to disorder, from low entropy to high entropy. The reverse can happen, but it is unlikely.

Macrostate		**Number of microstates**
Heads	Tails	(W)
100	0	1
99	1	1.0×10^{2}
95	5	7.5×10^{7}
90	10	1.7×10^{13}

Macrostate		Number of microstates
75	25	2.4×10^{23}
60	40	1.4×10^{28}
55	45	6.1×10^{28}
51	49	9.9×10^{28}
50	50	1.0×10^{29}
49	51	9.9×10^{28}
45	55	6.1×10^{28}
40	60	1.4×10^{28}
25	75	2.4×10^{23}

Macrostate		Number of microstates
10	90	1.7×10^{13}
5	95	7.5×10^{7}
1	99	1.0×10^{2}
0	100	1
		Total: 1.27×10^{30}

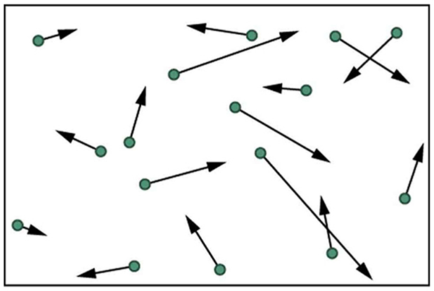
100-Coin Toss

This result becomes dramatic for larger systems. Consider what happens if you have 100 coins instead of just 5. The most orderly arrangements (most structured) are 100 heads or 100 tails. The least orderly (least structured) is that of 50 heads and 50 tails. There is only 1 way (1 microstate) to get the most orderly arrangement of 100 heads. There are 100 ways (100 microstates) to get the next most orderly arrangement of 99 heads and 1 tail (also 100 to get its reverse). And there are 1.0×10^{29} ways to get 50 heads and 50 tails, the least orderly arrangement. [link] is an abbreviated list of the various macrostates and the number of microstates for each macrostate. The total number of microstates—the total number of different ways 100 coins can be tossed—is an impressively large 1.27×10^{30}. Now, if we start with an orderly macrostate like 100 heads and toss the coins, there is a virtual certainty that we will get a less orderly macrostate. If we keep tossing the coins, it is possible, but exceedingly unlikely, that we will ever get back to the most orderly macrostate. If you tossed the coins once each second, you could expect to get either 100 heads or 100 tails once in 2×10^{22} years! This period is 1 trillion (10^{12}) times longer than the age of the universe, and so the chances are essentially zero. In contrast, there is an 8% chance of getting 50 heads, a 73% chance of getting from 45 to 55 heads, and a 96% chance of getting from 40 to 60 heads. Disorder is highly likely.

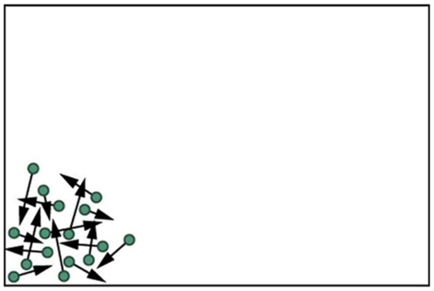
Disorder in a Gas

The fantastic growth in the odds favoring disorder that we see in going from 5 to 100 coins continues as the number of entities in the system increases. Let us now imagine applying this approach to perhaps a small sample of gas. Because counting microstates and macrostates involves statistics, this is called **statistical analysis**. The macrostates of a gas correspond to its macroscopic properties, such as volume, temperature, and pressure; and its microstates correspond to the detailed description of the positions and velocities of its atoms. Even a small amount of gas has a huge number of atoms: 1.0 cm^3 of an ideal gas at 1.0 atm and 0^{o} C has 2.7×10^{19} atoms. So each macrostate has an immense number of microstates. In plain language, this means that there are an immense number of ways in which the atoms in a gas can be arranged, while still having the same pressure, temperature, and so on.

The most likely conditions (or macrostates) for a gas are those we see all the time—a random distribution of atoms in space with a Maxwell-Boltzmann distribution of speeds in random directions, as predicted by kinetic theory. This is the most disorderly and least structured condition we can imagine. In contrast, one type of very orderly and structured macrostate has all of the atoms in one corner of a container with identical velocities. There are very few ways to accomplish this (very few microstates corresponding to it), and so it is exceedingly unlikely ever to occur. (See [link](b).) Indeed, it is so unlikely that we have a law saying that it is impossible, which has never been observed to be violated—the second law of thermodynamics.



(a) Likely



(b) Highly unlikely

(a) The ordinary state of gas in a container is a disorderly, random distribution of atoms or molecules with a Maxwell-Boltzmann distribution of speeds. It is so unlikely that these atoms or molecules would ever end up in one corner of the container that it might as well be impossible. (b) With energy transfer, the gas can be forced into one corner and its entropy greatly reduced. But left alone, it will spontaneously increase its entropy and return to the normal conditions, because they are immensely more likely.

The disordered condition is one of high entropy, and the ordered one has low entropy. With a transfer of energy from another system, we could force all of the atoms into one corner and have a local decrease in entropy, but at the cost of an overall increase in entropy of the universe. If the atoms start out in one corner, they will quickly disperse and become uniformly distributed and will never return to the orderly original state ([link](b)). Entropy will increase. With such a large sample of atoms, it is possible—but unimaginably unlikely—for entropy to decrease. Disorder is vastly more likely than order.

The arguments that disorder and high entropy are the most probable states are quite convincing. The great Austrian physicist Ludwig Boltzmann (1844–1906)—who, along with Maxwell, made so many contributions to kinetic theory—proved that the entropy of a system in a given state (a macrostate) can be written as

Equation:

$$S = k \ln W,$$

where $k = 1.38 \times 10^{-23}\ \mathrm{J/K}$ is Boltzmann's constant, and $\ln W$ is the natural logarithm of the number of microstates W corresponding to the given macrostate. W is proportional to the probability that the macrostate will occur. Thus entropy is directly related to the probability of a state—the more likely the state, the greater its entropy. Boltzmann proved that this expression for S is equivalent to the definition $\Delta S = Q/T$, which we have used extensively.

Thus the second law of thermodynamics is explained on a very basic level: entropy either remains the same or increases in every process. This phenomenon is due to the extraordinarily small probability of a decrease, based on the extraordinarily larger number of microstates in systems with greater entropy. Entropy *can* decrease, but for any macroscopic system, this outcome is so unlikely that it will never be observed.

Example:

Entropy Increases in a Coin Toss

Suppose you toss 100 coins starting with 60 heads and 40 tails, and you get the most likely result, 50 heads and 50 tails. What is the change in entropy?

Strategy

Noting that the number of microstates is labeled W in [link] for the 100-coin toss, we can use $\Delta S = S_\text{f} - S_\text{i} = k \ln W_\text{f} - k \ln W_\text{i}$ to calculate the change in entropy.

Solution

The change in entropy is

Equation:

$$\Delta S = S_\text{f} - S_\text{i} = k \ln W_\text{f} - k \ln W_\text{i},$$

where the subscript i stands for the initial 60 heads and 40 tails state, and the subscript f for the final 50 heads and 50 tails state. Substituting the values for W from [link] gives

Equation:

$$\begin{array}{lll}\Delta S & = & (1.38 \times 10^{-23}\ \text{J/K})[\ln(1.0 \times 10^{29}) - \ln(1.4 \times 10^{28})] \\ & = & 2.7 \times 10^{-23}\ \text{J/K}\end{array}$$

Discussion

This increase in entropy means we have moved to a less orderly situation. It is not impossible for further tosses to produce the initial state of 60 heads and 40 tails, but it is less likely. There is about a 1 in 90 chance for that decrease in entropy ($-2.7 \times 10^{-23}\ \text{J/K}$) to occur. If we calculate the decrease in entropy to move to the most orderly state, we get

$\Delta S = -92 \times 10^{-23}\ \text{J/K}$. There is about a 1 in 10^{30} chance of this change occurring. So while very small decreases in entropy are unlikely, slightly greater decreases are impossibly unlikely. These probabilities imply, again, that for a macroscopic system, a decrease in entropy is impossible. For example, for heat transfer to occur spontaneously from 1.00 kg of 0°C ice to its 0°C environment, there would be a decrease in entropy of $1.22 \times 10^{3}\ \text{J/K}$. Given that a ΔS of $10^{-21}\ \text{J/K}$ corresponds to about a 1 in 10^{30} chance, a decrease of this size ($10^{3}\ \text{J/K}$) is an *utter* impossibility. Even for a milligram of melted ice to spontaneously refreeze is impossible.

Note:

Problem-Solving Strategies for Entropy

1. *Examine the situation to determine if entropy is involved.*
2. *Identify the system of interest and draw a labeled diagram of the system showing energy flow.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns).* A written list is useful.
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns).* You must carefully identify the heat transfer, if any, and the temperature at which the process takes place. It is also important to identify the initial and final states.
5. *Solve the appropriate equation for the quantity to be determined (the unknown).* Note that the change in entropy can be determined between any states by calculating it for a reversible process.
6. *Substitute the known value along with their units into the appropriate equation, and obtain numerical solutions complete with units.*
7. *To see if it is reasonable: Does it make sense?* For example, total entropy should increase for any real process or be constant for a reversible process. Disordered states should be more probable and have greater entropy than ordered states.

Section Summary

- Disorder is far more likely than order, which can be seen statistically.
- The entropy of a system in a given state (a macrostate) can be written as
  Equation:
  $$S = k \ln W,$$
  where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, and $\ln W$ is the natural logarithm of the number of microstates W corresponding to the given macrostate.

Conceptual Questions

Exercise:

Problem:

Explain why a building made of bricks has smaller entropy than the same bricks in a disorganized pile. Do this by considering the number of ways that each could be formed (the number of microstates in each macrostate).

Problem Exercises

Exercise:

Problem:

Using [link], verify the contention that if you toss 100 coins each second, you can expect to get 100 heads or 100 tails once in 2×10^{22} years; calculate the time to two-digit accuracy.

Solution:

It should happen twice in every 1.27×10^{30} s or once in every 6.35×10^{29} s

$$6.35 \times 10^{29}\ \text{s} \left(\frac{1\ \text{h}}{3600\ \text{s}}\right)\left(\frac{1\ \text{d}}{24\ \text{h}}\right)\left(\frac{1\ \text{y}}{365.25\ \text{d}}\right) = 2.0 \times 10^{22}\ \text{y}$$

Exercise:

Problem:

What percent of the time will you get something in the range from 60 heads and 40 tails through 40 heads and 60 tails when tossing 100 coins? The total number of microstates in that range is 1.22×10^{30}. (Consult [link].)

Exercise:

Problem:

(a) If tossing 100 coins, how many ways (microstates) are there to get the three most likely macrostates of 49 heads and 51 tails, 50 heads and 50 tails, and 51 heads and 49 tails? (b) What percent of the total possibilities is this? (Consult [link].)

Solution:

(a) 3.0×10^{29}

(b) 24%

Exercise:

Problem:

(a) What is the change in entropy if you start with 100 coins in the 45 heads and 55 tails macrostate, toss them, and get 51 heads and 49 tails? (b) What if you get 75 heads and 25 tails? (c) How much more likely is 51 heads and 49 tails than 75 heads and 25 tails? (d) Does either outcome violate the second law of thermodynamics?

Exercise:

Problem:

(a) What is the change in entropy if you start with 10 coins in the 5 heads and 5 tails macrostate, toss them, and get 2 heads and 8 tails? (b) How much more likely is 5 heads and 5 tails than 2 heads and 8 tails? (Take the ratio of the number of microstates to find out.) (c) If you were betting on 2 heads and 8 tails would you accept odds of 252 to 45? Explain why or why not.

Solution:

(a) $-2.38\times10^{-23}\ \text{J/K}$

(b) 5.6 times more likely

(c) If you were betting on two heads and 8 tails, the odds of breaking even are 252 to 45, so on average you would break even. So, no, you wouldn’t bet on odds of 252 to 45.

Macrostate		**Number of Microstates (*W*)**
Heads	Tails	
10	0	1
9	1	10
8	2	45
7	3	120

Macrostate		**Number of Microstates (*W*)**
6	4	210
5	5	252
4	6	210
3	7	120
2	8	45
1	9	10
0	10	1
		Total: 1024

10-Coin Toss

Exercise:

Problem:

(a) If you toss 10 coins, what percent of the time will you get the three most likely macrostates (6 heads and 4 tails, 5 heads and 5 tails, 4 heads and 6 tails)? (b) You can realistically toss 10 coins and count the number of heads and tails about twice a minute. At that rate, how long will it take on average to get either 10 heads and 0 tails or 0 heads and 10 tails?

Exercise:

Problem:

(a) Construct a table showing the macrostates and all of the individual microstates for tossing 6 coins. (Use [link] as a guide.) (b) How many macrostates are there? (c) What is the total number of microstates? (d) What percent chance is there of tossing 5 heads and 1 tail? (e) How much more likely are you to toss 3 heads and 3 tails than 5 heads and 1 tail? (Take the ratio of the number of microstates to find out.)

Solution:

(b) 7

(c) 64

(d) 9.38%

(e) 3.33 times more likely (20 to 6)

Exercise:

Problem:

In an air conditioner, 12.65 MJ of heat transfer occurs from a cold environment in 1.00 h. (a) What mass of ice melting would involve the same heat transfer? (b) How many hours of operation would be equivalent to melting 900 kg of ice? (c) If ice costs 20 cents per kg, do you think the air conditioner could be operated more cheaply than by simply using ice? Describe in detail how you evaluate the relative costs.

Glossary

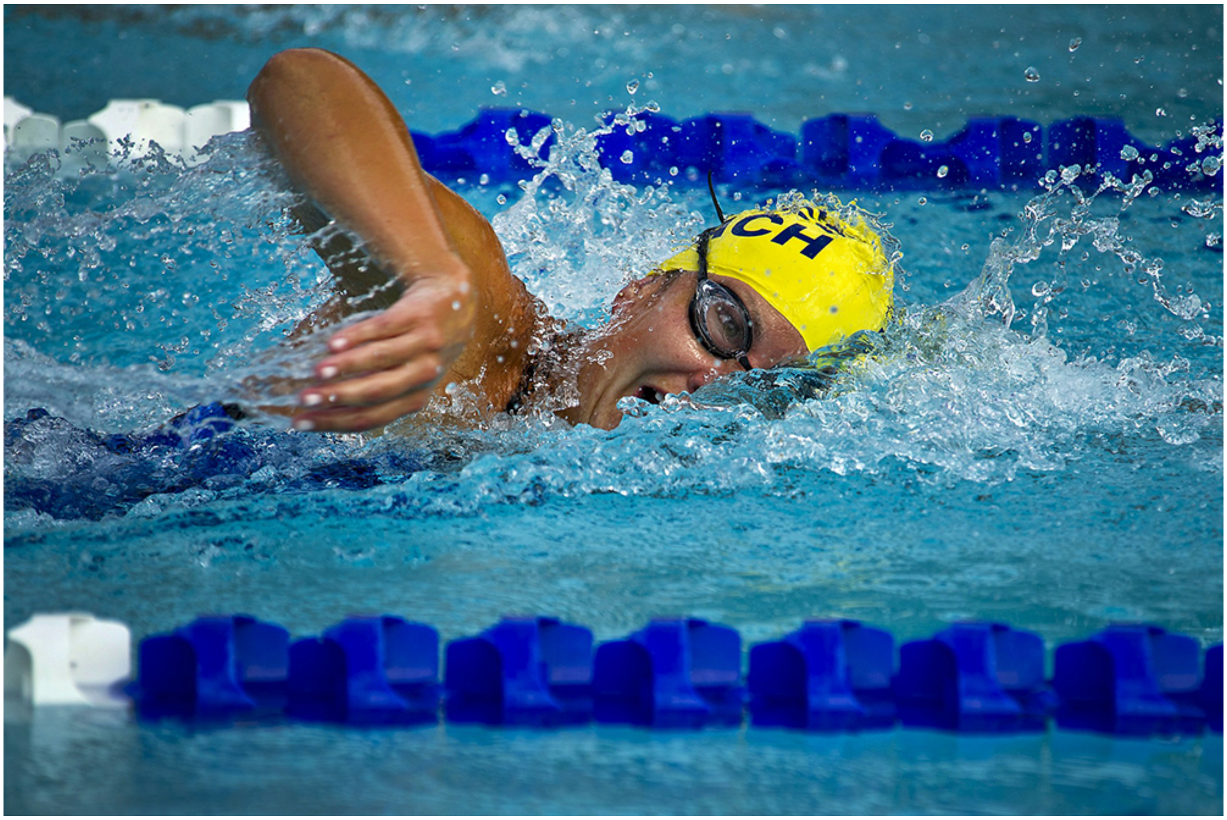
macrostate	an overall property of a system
microstate	each sequence within a larger macrostate

statistical analysis

using statistics to examine data, such as counting microstates and macrostates

Introduction to Fluid Statics class="introduction"

The fluid essential to all life has a beauty of its own. It also helps support the weight of this swimmer. (credit: 12019, Pixabay)

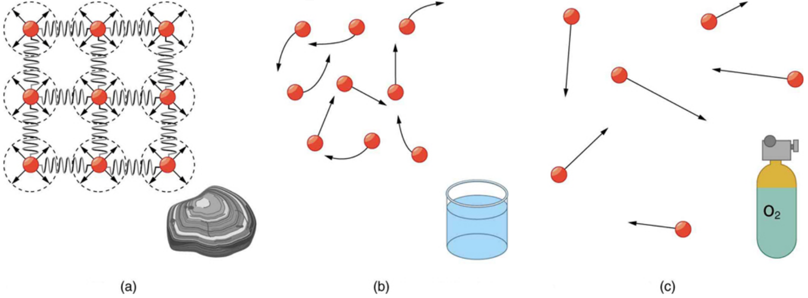


Much of what we value in life is fluid: a breath of fresh winter air; the hot blue flame in our gas cooker; the water we drink, swim in, and bathe in; the blood in our veins. What exactly is a fluid? Can we understand fluids with the laws already presented, or will new laws emerge from their study? The physical characteristics of static or stationary fluids and some of the laws that govern their behavior are the topics of this chapter. Fluid Dynamics and Its Biological and Medical Applications explores aspects of fluid flow.

What Is a Fluid?

- State the common phases of matter.
- Explain the physical characteristics of solids, liquids, and gases.
- Describe the arrangement of atoms in solids, liquids, and gases.

Matter most commonly exists as a solid, liquid, or gas; these states are known as the three common *phases of matter*. Solids have a definite shape and a specific volume, liquids have a definite volume but their shape changes depending on the container in which they are held, and gases have neither a definite shape nor a specific volume as their molecules move to fill the container in which they are held. (See [link].) Liquids and gases are considered to be fluids because they yield to shearing forces, whereas solids resist them. Note that the extent to which fluids yield to shearing forces (and hence flow easily and quickly) depends on a quantity called the viscosity which is discussed in detail in Viscosity and Laminar Flow; Poiseuille's Law. We can understand the phases of matter and what constitutes a fluid by considering the forces between atoms that make up matter in the three phases.



(a) Atoms in a solid always have the same neighbors, held near home by forces represented here by springs. These atoms are essentially in contact with one another. A rock is an example of a solid. This rock retains its shape because of the forces holding its atoms together. (b) Atoms in a liquid are also in close contact but can slide over one another. Forces between them strongly resist attempts to push them closer together and also hold them in close contact.

Water is an example of a liquid. Water can flow, but it also remains in an open container because of the forces between its atoms. (c) Atoms in a gas are separated by distances that are considerably larger than the size of the atoms themselves, and they move about freely. A gas must be held in a closed container to prevent it from moving out freely.

Atoms in *solids* are in close contact, with forces between them that allow the atoms to vibrate but not to change positions with neighboring atoms. (These forces can be thought of as springs that can be stretched or compressed, but not easily broken.) Thus a solid *resists* all types of stress. A solid cannot be easily deformed because the atoms that make up the solid are not able to move about freely. Solids also resist compression, because their atoms form part of a lattice structure in which the atoms are a relatively fixed distance apart. Under compression, the atoms would be forced into one another. Most of the examples we have studied so far have involved solid objects which deform very little when stressed.

Note:

Connections: Submicroscopic Explanation of Solids and Liquids

Atomic and molecular characteristics explain and underlie the macroscopic characteristics of solids and fluids. This submicroscopic explanation is one theme of this text and is highlighted in the Things Great and Small features in Conservation of Momentum. See, for example, microscopic description of collisions and momentum or microscopic description of pressure in a gas. This present section is devoted entirely to the submicroscopic explanation of solids and liquids.

In contrast, *liquids* deform easily when stressed and do not spring back to their original shape once the force is removed because the atoms are free to slide about and change neighbors—that is, they *flow* (so they are a type of fluid), with the molecules held together by their mutual attraction. When a liquid is placed in a container with no lid on, it remains in the container (providing the container has no holes below the surface of the liquid!). Because the atoms are closely packed, liquids, like solids, resist compression.

Atoms in *gases* are separated by distances that are large compared with the size of the atoms. The forces between gas atoms are therefore very weak, except when the atoms collide with one another. Gases thus not only flow (and are therefore considered to be fluids) but they are relatively easy to compress because there is much space and little force between atoms. When placed in an open container gases, unlike liquids, will escape. The major distinction is that gases are easily compressed, whereas liquids are not. We shall generally refer to both gases and liquids simply as **fluids**, and make a distinction between them only when they behave differently.

Note:
PhET Explorations: States of Matter—Basics
Heat, cool, and compress atoms and molecules and watch as they change between solid, liquid, and gas phases.
https://phet.colorado.edu/sims/html/states-of-matter-basics/latest/states-of-matter-basics_en.html

Section Summary

- A fluid is a state of matter that yields to sideways or shearing forces. Liquids and gases are both fluids. Fluid statics is the physics of stationary fluids.

Conceptual Questions

Exercise:

Problem:

What physical characteristic distinguishes a fluid from a solid?

Exercise:

Problem:

Which of the following substances are fluids at room temperature: air, mercury, water, glass?

Exercise:

Problem: Why are gases easier to compress than liquids and solids?

Exercise:

Problem: How do gases differ from liquids?

Glossary

fluids	liquids and gases; a fluid is a state of matter that yields to shearing forces

Density

- Define density.
- Calculate the mass of a reservoir from its density.
- Compare and contrast the densities of various substances.

Which weighs more, a ton of feathers or a ton of bricks? This old riddle plays with the distinction between mass and density. A ton is a ton, of course; but bricks have much greater density than feathers, and so we are tempted to think of them as heavier. (See [link].)

Density, as you will see, is an important characteristic of substances. It is crucial, for example, in determining whether an object sinks or floats in a fluid. Density is the mass per unit volume of a substance or object. In equation form, density is defined as

Equation:

$$\rho = \frac{m}{V},$$

where the Greek letter ρ (rho) is the symbol for density, m is the mass, and V is the volume occupied by the substance.

Note:

Density

Density is mass per unit volume.

Equation:

$$\rho = \frac{m}{V},$$

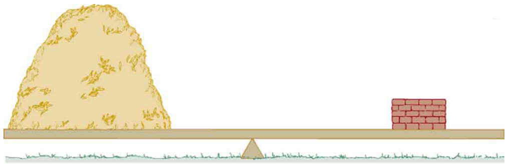
where ρ is the symbol for density, m is the mass, and V is the volume occupied by the substance.

In the riddle regarding the feathers and bricks, the masses are the same, but the volume occupied by the feathers is much greater, since their density is much lower. The SI unit of density is kg/m^3, representative values are given in [link]. The metric system was originally devised so that water would have a density of $1\ \text{g/cm}^3$, equivalent to $10^3\ \text{kg/m}^3$. Thus the basic mass unit, the kilogram, was first devised to be the mass of 1000 mL of water, which has a volume of $1000\ \text{cm}^3$.

Substance	$\rho (10^3\ \text{kg/m}^3\ \text{or g/mL})$	**Substance**	$\rho (10^3\ \text{kg/m}^3\ \text{or g/mL})$	**Substance**	$\rho($
Solids		**Liquids**		**Gases**	
Aluminum	2.7	Water (4ºC)	1.000	Air	

Substance	$\rho(10^3\ \mathrm{kg/m^3}\ \text{or}\ \mathrm{g/mL})$	**Substance**	$\rho(10^3\ \mathrm{kg/m^3}\ \text{or}\ \mathrm{g/mL})$	**Substance**	$\rho($
Brass	8.44	Blood	1.05	Carbon dioxide	
Copper (average)	8.8	Sea water	1.025	Carbon monoxide	
Gold	19.32	Mercury	13.6	Hydrogen	
Iron or steel	7.8	Ethyl alcohol	0.79	Helium	
Lead	11.3	Petrol	0.68	Methane	
Polystyrene	0.10	Glycerin	1.26	Nitrogen	
Tungsten	19.30	Olive oil	0.92	Nitrous oxide	
Uranium	18.70			Oxygen	
Concrete	2.30–3.0			Steam $(100^o\ \mathrm{C})$	
Cork	0.24				
Glass, common (average)	2.6				
Granite	2.7				
Earth’s crust	3.3				
Wood	0.3–0.9				
Ice (0°C)	0.917				
Bone	1.7–2.0				

Densities of Various Substances



A ton of feathers and a ton of bricks have the same mass, but the feathers make a much bigger pile because they have a much lower density.

As you can see by examining [link], the density of an object may help identify its composition. The density of gold, for example, is about 2.5 times the density of iron, which is about 2.5 times the density of aluminum. Density also reveals something about the phase of the matter and its substructure. Notice that the densities of liquids and solids are roughly comparable, consistent with the fact that their atoms are in close contact. The densities of gases are much less than those of liquids and solids, because the atoms in gases are separated by large amounts of empty space.

Note:
Take-Home Experiment Sugar and Salt
A pile of sugar and a pile of salt look pretty similar, but which weighs more? If the volumes of both piles are the same, any difference in mass is due to their different densities (including the air space between crystals). Which do you think has the greater density? What values did you find? What method did you use to determine these values?

Example:
Calculating the Mass of a Reservoir From Its Volume
A reservoir has a surface area of 50.0 km^2 and an average depth of 40.0 m. What mass of water is held behind the dam? (See [link] for a view of a large reservoir—the Three Gorges Dam site on the Yangtze River in central China.)
Strategy
We can calculate the volume V of the reservoir from its dimensions, and find the density of water ρ in [link]. Then the mass m can be found from the definition of density
Equation:

$$\rho = \frac{m}{V}.$$

Solution
Solving equation $\rho = m/V$ for m gives $m = \rho V$.
The volume V of the reservoir is its surface area A times its average depth h:
Equation:

$$\begin{aligned} V &= \text{Ah} = \quad 50.0\text{ km}^2 \ (40.0\text{ m}) \\ &= \quad 50.0\text{ km}^2 \ \ \frac{10^3\text{ m}}{1\text{ km}}^{\ 2} \ (40.0\text{ m}) = 2.00 \times 10^9\text{ m}^3 \end{aligned}$$

The density of water ρ from [link] is $1.000 \times 10^3\text{ kg/m}^3$. Substituting V and ρ into the expression for mass gives
Equation:

$$\begin{aligned} m &= 1.00 \times 10^{3}\ \mathrm{kg/m}^{3} \quad 2.00 \times 10^{9}\ \mathrm{m}^{3} \\ &= 2.00 \times 10^{12}\ \mathrm{kg}. \end{aligned}$$

Discussion

A large reservoir contains a very large mass of water. In this example, the weight of the water in the reservoir is $\mathrm{mg} = 1.96 \times 10^{13}\ \mathrm{N}$, where *g* is the acceleration due to the Earth's gravity (about $9.80\ \mathrm{m/s}^{2}$). It is reasonable to ask whether the dam must supply a force equal to this tremendous weight. The answer is no. As we shall see in the following sections, the force the dam must supply can be much smaller than the weight of the water it holds back.

Three Gorges Dam in central China. When completed in 2008, this became the world's largest hydroelectric plant, generating power equivalent to that generated by 22 average-sized nuclear power plants. The concrete dam is 181 m high and 2.3 km across. The reservoir made by this dam is 660 km long. Over 1 million people were displaced by the creation of the reservoir. (credit: Le Grand Portage)

Section Summary

- Density is the mass per unit volume of a substance or object. In equation form, density is defined as
  Equation:

$$\rho = \frac{m}{V}.$$

- The SI unit of density is $\mathrm{kg/m}^{3}$.

Conceptual Questions

Exercise:

Problem: Approximately how does the density of air vary with altitude?

Exercise:

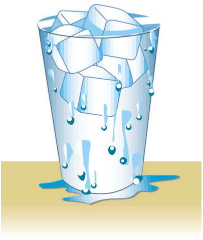
Problem:

Give an example in which density is used to identify the substance composing an object. Would information in addition to average density be needed to identify the substances in an object composed of more than one material?

Exercise:

Problem:

[link] shows a glass of ice water filled to the brim. Will the water overflow when the ice melts? Explain your answer.



Problems & Exercises

Exercise:

Problem: Gold is sold by the troy ounce (31.103 g). What is the volume of 1 troy ounce of pure gold?

Solution:

1.610 cm^3

Exercise:

Problem:

Mercury is commonly supplied in flasks containing 34.5 kg (about 76 lb). What is the volume in liters of this much mercury?

Exercise:

Problem:

(a) What is the mass of a deep breath of air having a volume of 2.00 L? (b) Discuss the effect taking such a breath has on your body's volume and density.

Solution:

(a) 2.58 g

(b) The volume of your body increases by the volume of air you inhale. The average density of your body decreases when you take a deep breath, because the density of air is substantially smaller than the average density of the body before you took the deep breath.

Exercise:

Problem:

A straightforward method of finding the density of an object is to measure its mass and then measure its volume by submerging it in a graduated cylinder. What is the density of a 240-g rock that displaces 89.0 cm^3 of water? (Note that the accuracy and practical applications of this technique are more limited than a variety of others that are based on Archimedes' principle.)

Solution:

2.70 g/cm^3

Exercise:

Problem:

Suppose you have a coffee mug with a circular cross section and vertical sides (uniform radius). What is its inside radius if it holds 375 g of coffee when filled to a depth of 7.50 cm? Assume coffee has the same density as water.

Exercise:

Problem:

(a) A rectangular gasoline tank can hold 50.0 kg of gasoline when full. What is the depth of the tank if it is 0.500-m wide by 0.900-m long? (b) Discuss whether this gas tank has a reasonable volume for a passenger car.

Solution:

(a) 0.163 m

(b) Equivalent to 19.4 gallons, which is reasonable

Exercise:

Problem:

A trash compactor can reduce the volume of its contents to 0.350 their original value. Neglecting the mass of air expelled, by what factor is the density of the rubbish increased?

Exercise:

Problem:

A 2.50-kg steel gasoline can holds 20.0 L of gasoline when full. What is the average density of the full gas can, taking into account the volume occupied by steel as well as by gasoline?

Solution:

$7.9 \times 10^2\text{ kg/m}^3$

Exercise:

Problem:

What is the density of 18.0-karat gold that is a mixture of 18 parts gold, 5 parts silver, and 1 part copper? (These values are parts by mass, not volume.) Assume that this is a simple mixture having an average density equal to the weighted densities of its constituents.

Solution:

15.6 g/cm^3

Exercise:

Problem:

There is relatively little empty space between atoms in solids and liquids, so that the average density of an atom is about the same as matter on a macroscopic scale—approximately $10^3\ \text{kg/m}^3$. The nucleus of an atom has a radius about 10^{-5} that of the atom and contains nearly all the mass of the entire atom. (a) What is the approximate density of a nucleus? (b) One remnant of a supernova, called a neutron star, can have the density of a nucleus. What would be the radius of a neutron star with a mass 10 times that of our Sun (the radius of the Sun is $7 \times 10^8\ \text{m}$)?

Solution:

(a) $10^{18}\ \text{kg/m}^3$

(b) $2 \times 10^4\ \text{m}$

Glossary

density
: the mass per unit volume of a substance or object

Pressure

- Define pressure.
- Explain the relationship between pressure and force.
- Calculate force given pressure and area.

You have no doubt heard the word **pressure** being used in relation to blood (high or low blood pressure) and in relation to the weather (high- and low-pressure weather systems). These are only two of many examples of pressures in fluids. Pressure P is defined as

Equation:

$$P=\frac{F}{A}$$

where F is a force applied to an area A that is perpendicular to the force.

Note:

Pressure

Pressure is defined as the force divided by the area perpendicular to the force over which the force is applied, or

Equation:

$$P=\frac{F}{A}.$$

A given force can have a significantly different effect depending on the area over which the force is exerted, as shown in [link]. The SI unit for pressure is the *pascal*, where

Equation:

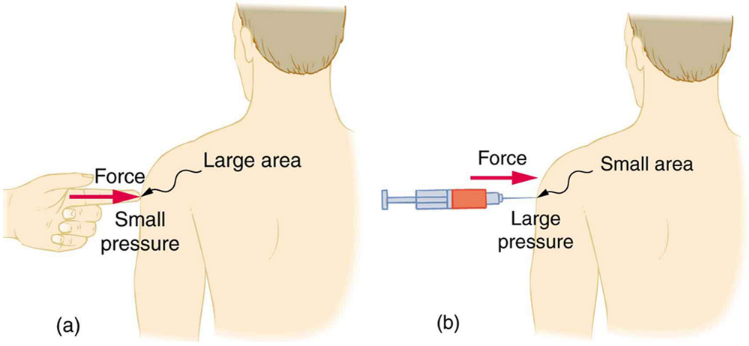
$$1\text{ Pa}=1\text{ N/m}^2.$$

In addition to the pascal, there are many other units for pressure that are in common use. In meteorology, atmospheric pressure is often described in units of millibar (mb), where

Equation:

$$100\text{ mb}=1\times 10^{4}\text{ Pa}\ \ .$$

Pounds per square inch $\left(\text{lb/in}^{2}\text{ or psi}\right)$ is still sometimes used as a measure of tire pressure, and millimeters of mercury (mm Hg) is still often used in the measurement of blood pressure. Pressure is defined for all states of matter but is particularly important when discussing fluids.



(a) While the person being poked with the finger might be irritated, the force has little lasting effect. (b) In contrast, the same force applied to an area the size of the sharp end of a needle is great enough to break the skin.

Example:

Calculating Force Exerted by the Air: What Force Does a Pressure Exert?

An astronaut is working outside the International Space Station where the atmospheric pressure is essentially zero. The pressure gauge on her air tank reads 6.90×10^6 Pa. What force does the air inside the tank exert on the flat end of the cylindrical tank, a disk 0.150 m in diameter?

Strategy

We can find the force exerted from the definition of pressure given in $P = \frac{F}{A}$, provided we can find the area A acted upon.

Solution

By rearranging the definition of pressure to solve for force, we see that

Equation:

$$F = \text{PA}.$$

Here, the pressure P is given, as is the area of the end of the cylinder A, given by $A = \pi r^2$. Thus,

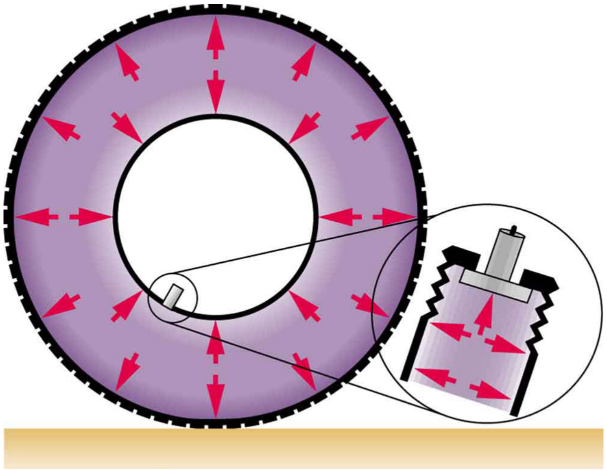
Equation:

$$\begin{aligned} F &= \left(6.90 \times 10^6\ \text{N/m}^2\right)(3.14)(0.0750\ \text{m})^2 \\ &= 1.22 \times 10^5\ \text{N}. \end{aligned}$$

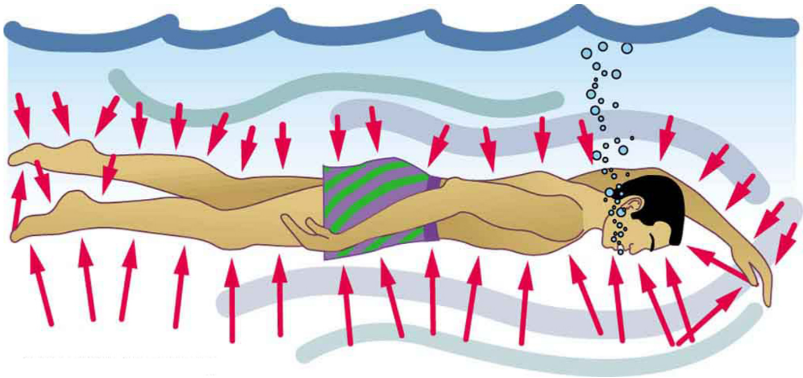
Discussion

Wow! No wonder the tank must be strong. Since we found $F = \text{PA}$, we see that the force exerted by a pressure is directly proportional to the area acted upon as well as the pressure itself.

The force exerted on the end of the tank is perpendicular to its inside surface. This direction is because the force is exerted by a static or stationary fluid. We have already seen that fluids cannot *withstand* shearing (sideways) forces; they cannot *exert* shearing forces, either. Fluid pressure has no direction, being a scalar quantity. The forces due to pressure have well-defined directions: they are always exerted perpendicular to any surface. (See the tire in [link], for example.) Finally, note that pressure is exerted on all surfaces. Swimmers, as well as the tire, feel pressure on all sides. (See [link].)



Pressure inside this tire exerts forces perpendicular to all surfaces it contacts. The arrows give representative directions and magnitudes of the forces exerted at various points. Note that static fluids do not exert shearing forces.



Pressure is exerted on all sides of this swimmer, since the water would flow into the space he occupies if he were not there. The arrows represent the directions and magnitudes of the forces exerted at various points on the swimmer. Note that the forces are larger underneath, due to greater depth, giving a net upward or buoyant force that is balanced by the weight of the swimmer.

Note:
PhET Explorations: Gas Properties
Pump gas molecules to a box and see what happens as you change the volume, add or remove heat, change gravity, and more. Measure the temperature and pressure, and discover how the properties of the gas vary in relation to each other.

Gas Properties

Section Summary

- Pressure is the force per unit perpendicular area over which the force is applied. In equation form, pressure is defined as
  Equation:
  $$P = \frac{F}{A}.$$
- The SI unit of pressure is pascal and $1\ \text{Pa} = 1\ \text{N/m}^2$.

Conceptual Questions

Exercise:

Problem:

How is pressure related to the sharpness of a knife and its ability to cut?

Exercise:

Problem:

Why does a dull hypodermic needle hurt more than a sharp one?

Exercise:

Problem:

The outward force on one end of an air tank was calculated in [link]. How is this force balanced? (The tank does not accelerate, so the force must be balanced.)

Exercise:

Problem:

Why is force exerted by static fluids always perpendicular to a surface?

Exercise:

Problem:

In a remote location near the North Pole, an iceberg floats in a lake. Next to the lake (assume it is not frozen) sits a comparably sized glacier sitting on land. If both chunks of ice should melt due to rising global temperatures (and the melted ice all goes into the lake), which ice chunk would give the greatest increase in the level of the lake water, if any?

Exercise:

Problem:

How do jogging on soft ground and wearing padded shoes reduce the pressures to which the feet and legs are subjected?

Exercise:

Problem:

Toe dancing (as in ballet) is much harder on toes than normal dancing or walking. Explain in terms of pressure.

Exercise:

Problem:

How do you convert pressure units like millimeters of mercury, centimeters of water, and inches of mercury into units like newtons per meter squared without resorting to a table of pressure conversion factors?

Problems & Exercises

Exercise:

Problem:

As a woman walks, her entire weight is momentarily placed on one heel of her high-heeled shoes. Calculate the pressure exerted on the floor by the heel if it has an area of $1.50\ \text{cm}^2$ and the woman's mass is 55.0 kg. Express the pressure in Pa. (In the early days of commercial flight, women were not allowed to wear high-heeled shoes because aircraft floors were too thin to withstand such large pressures.)

Solution:

$3.59 \times 10^6\ \text{Pa}$; or $521\ \text{lb/in}^2$

Exercise:

Problem:

The pressure exerted by a phonograph needle on a record is surprisingly large. If the equivalent of 1.00 g is supported by a needle, the tip of which is a circle 0.200 mm in radius, what pressure is exerted on the record in $\mathrm{N/m^2}$?

Exercise:

Problem:

Nail tips exert tremendous pressures when they are hit by hammers because they exert a large force over a small area. What force must be exerted on a nail with a circular tip of 1.00 mm diameter to create a pressure of $3.00 \times 10^9\ \mathrm{N/m^2}$?(This high pressure is possible because the hammer striking the nail is brought to rest in such a short distance.)

Solution:

$2.36 \times 10^3\ \mathrm{N}$

Glossary

pressure	the force per unit area perpendicular to the force, over which the force acts

Variation of Pressure with Depth in a Fluid

- Define pressure in terms of weight.
- Explain the variation of pressure with depth in a fluid.
- Calculate density given pressure and altitude.

If your ears have ever popped on a plane flight or ached during a deep dive in a swimming pool, you have experienced the effect of depth on pressure in a fluid. At the Earth's surface, the air pressure exerted on you is a result of the weight of air above you. This pressure is reduced as you climb up in altitude and the weight of air above you decreases. Under water, the pressure exerted on you increases with increasing depth. In this case, the pressure being exerted upon you is a result of both the weight of water above you *and* that of the atmosphere above you. You may notice an air pressure change on an elevator ride that transports you many stories, but you need only dive a meter or so below the surface of a pool to feel a pressure increase. The difference is that water is much denser than air, about 775 times as dense.

Consider the container in [link]. Its bottom supports the weight of the fluid in it. Let us calculate the pressure exerted on the bottom by the weight of the fluid. That **pressure** is the weight of the fluid mg divided by the area A supporting it (the area of the bottom of the container):

Equation:

$$P=\frac{\text{mg}}{A}.$$

We can find the mass of the fluid from its volume and density:

Equation:

$$m=\rho V.$$

The volume of the fluid V is related to the dimensions of the container. It is

Equation:

$$V=\text{Ah},$$

where A is the cross-sectional area and h is the depth. Combining the last two equations gives

Equation:

$$m = \rho \text{Ah}.$$

If we enter this into the expression for pressure, we obtain

Equation:

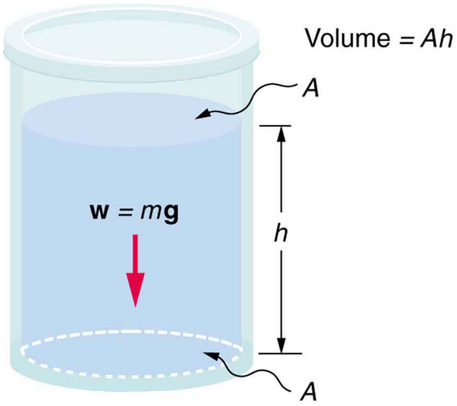
$$P = \frac{(\rho \text{Ah})g}{A}.$$

The area cancels, and rearranging the variables yields

Equation:

$$P = h\rho g.$$

This value is the *pressure due to the weight of a fluid*. The equation has general validity beyond the special conditions under which it is derived here. Even if the container were not there, the surrounding fluid would still exert this pressure, keeping the fluid static. Thus the equation $P = h\rho g$ represents the pressure due to the weight of any fluid of *average density* ρ at any depth h below its surface. For liquids, which are nearly incompressible, this equation holds to great depths. For gases, which are quite compressible, one can apply this equation as long as the density changes are small over the depth considered. [link] illustrates this situation.



The bottom of this container supports the entire weight of the fluid in it. The vertical sides cannot exert an upward force on the fluid (since it cannot withstand a shearing force), and so the bottom must support it all.

Example:

Calculating the Average Pressure and Force Exerted: What Force Must a Dam Withstand?

In [link], we calculated the mass of water in a large reservoir. We will now consider the pressure and force acting on the dam retaining water. (See [link].) The dam is 500 m wide, and the water is 80.0 m deep at the dam. (a) What is the average pressure on the dam due to the water? (b) Calculate the force exerted against the dam and compare it with the weight of water in the dam (previously found to be 1.96×10^{13} N).

Strategy for (a)

The average pressure P due to the weight of the water is the pressure at the average depth h of 40.0 m, since pressure increases linearly with depth.

Solution for (a)

The average pressure due to the weight of a fluid is

Equation:

$$P = h\rho g.$$

Entering the density of water from [link] and taking h to be the average depth of 40.0 m, we obtain

Equation:

$$\begin{aligned} P &= (40.0\ \text{m})\left(10^{3}\ \frac{\text{kg}}{\text{m}^{3}}\right)\left(9.80\ \frac{\text{m}}{\text{s}^{2}}\right) \\ &= 3.92 \times 10^{5}\ \frac{\text{N}}{\text{m}^{2}} = 392\ \text{kPa}. \end{aligned}$$

Strategy for (b)

The force exerted on the dam by the water is the average pressure times the area of contact:

Equation:

$$F = PA.$$

Solution for (b)

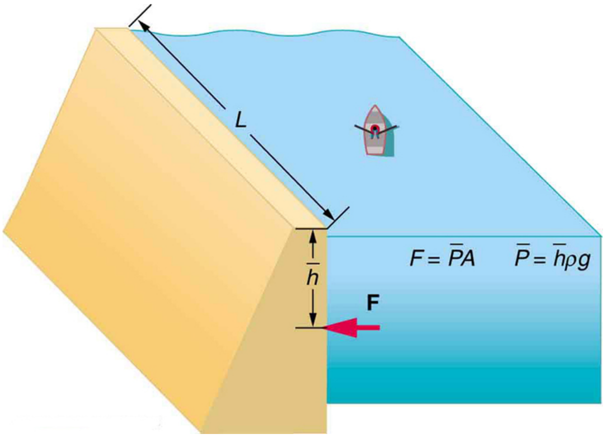
We have already found the value for P. The area of the dam is $A = 80.0\ \text{m} \times 500\ \text{m} = 4.00 \times 10^{4}\ \text{m}^{2}$, so that

Equation:

$$\begin{aligned} F &= (3.92 \times 10^{5}\ \text{N/m}^{2})(4.00 \times 10^{4}\ \text{m}^{2}) \\ &= 1.57 \times 10^{10}\ \text{N}. \end{aligned}$$

Discussion

Although this force seems large, it is small compared with the $1.96 \times 10^{13}\ \text{N}$ weight of the water in the reservoir—in fact, it is only 0.0800% of the weight. Note that the pressure found in part (a) is completely independent of the width and length of the lake—it depends only on its average depth at the dam. Thus the force depends only on the water’s average depth and the dimensions of the dam, *not* on the horizontal extent of the reservoir. In the diagram, the thickness of the dam increases with depth to balance the increasing force due to the increasing pressure.epth to balance the increasing force due to the increasing pressure.



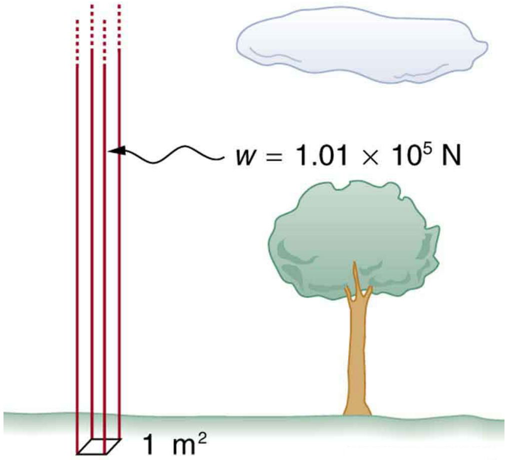
The dam must withstand the force exerted against it by the water it retains. This force is small compared with the weight of the water behind the dam.

Atmospheric pressure is another example of pressure due to the weight of a fluid, in this case due to the weight of *air* above a given height. The atmospheric pressure at the Earth’s surface varies a little due to the large-scale flow of the atmosphere induced by the Earth’s rotation (this creates weather “highs” and “lows”). However, the average pressure at sea level is given by the *standard atmospheric pressure* P_{atm}, measured to be

Equation:

$$1\ \text{atmosphere (atm)} = P_{\text{atm}} = 1.01 \times 10^5\ \text{N/m}^2 = 101\ \text{kPa}.$$

This relationship means that, on average, at sea level, a column of air above $1.00\ \text{m}^2$ of the Earth's surface has a weight of $1.01 \times 10^5\ \text{N}$, equivalent to $1\ \text{atm}$. (See [link].)



Atmospheric pressure at sea level averages $1.01 \times 10^5\ \text{Pa}$ (equivalent to 1 atm), since the column of air over this $1\ \text{m}^2$, extending to the top of the atmosphere, weighs $1.01 \times 10^5\ \text{N}$.

Example:

Calculating Average Density: How Dense Is the Air?

Calculate the average density of the atmosphere, given that it extends to an altitude of 120 km. Compare this density with that of air listed in [link].

Strategy

If we solve $P = h\rho g$ for density, we see that

Equation:

$$\rho = \frac{P}{\text{hg}}.$$

We then take P to be atmospheric pressure, h is given, and g is known, and so we can use this to calculate ρ.

Solution

Entering known values into the expression for ρ yields

Equation:

$$\rho = \frac{1.01 \times 10^5 \text{ N/m}^2}{(120 \times 10^3 \text{ m})(9.80 \text{ m/s}^2)} = 8.59 \times 10^{-2} \text{ kg/m}^3.$$

Discussion

This result is the average density of air between the Earth’s surface and the top of the Earth’s atmosphere, which essentially ends at 120 km. The density of air at sea level is given in [link] as 1.29 kg/m^3 —about 15 times its average value. Because air is so compressible, its density has its highest value near the Earth’s surface and declines rapidly with altitude.

Example:

Calculating Depth Below the Surface of Water: What Depth of Water Creates the Same Pressure as the Entire Atmosphere?

Calculate the depth below the surface of water at which the pressure due to the weight of the water equals 1.00 atm.

Strategy

We begin by solving the equation $P = h\rho g$ for depth h:

Equation:

$$h = \frac{P}{\rho g}.$$

Then we take P to be 1.00 atm and ρ to be the density of the water that creates the pressure.

Solution

Entering the known values into the expression for h gives

Equation:

$$h=\frac{1.01\times10^{5}\ \mathrm{N/m^{2}}}{(1.00\times10^{3}\ \mathrm{kg/m^{3}})(9.80\ \mathrm{m/s^{2}})}=10.3\ \mathrm{m}.$$

Discussion
Just 10.3 m of water creates the same pressure as 120 km of air. Since water is nearly incompressible, we can neglect any change in its density over this depth.

What do you suppose is the *total* pressure at a depth of 10.3 m in a swimming pool? Does the atmospheric pressure on the water's surface affect the pressure below? The answer is yes. This seems only logical, since both the water's weight and the atmosphere's weight must be supported. So the *total* pressure at a depth of 10.3 m is 2 atm—half from the water above and half from the air above. We shall see in Pascal's Principle that fluid pressures always add in this way.

Section Summary

- Pressure is the weight of the fluid mg divided by the area A supporting it (the area of the bottom of the container):
  Equation:

  $$P=\frac{\mathrm{mg}}{A}.$$

- Pressure due to the weight of a liquid is given by
  Equation:

  $$P=h\rho g,$$

  where P is the pressure, h is the height of the liquid, ρ is the density of the liquid, and g is the acceleration due to gravity.

Conceptual Questions

Exercise:

Problem:

Atmospheric pressure exerts a large force (equal to the weight of the atmosphere above your body—about 10 tons) on the top of your body when you are lying on the beach sunbathing. Why are you able to get up?

Exercise:

Problem:

Why does atmospheric pressure decrease more rapidly than linearly with altitude?

Exercise:

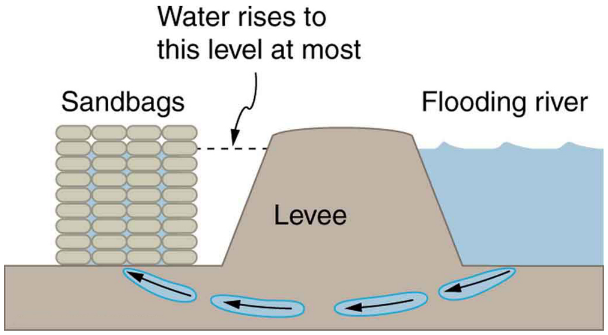
Problem:

What are two reasons why mercury rather than water is used in barometers?

Exercise:

Problem:

[link] shows how sandbags placed around a leak outside a river levee can effectively stop the flow of water under the levee. Explain how the small amount of water inside the column formed by the sandbags is able to balance the much larger body of water behind the levee.



Because the river level is very high, it has started to leak under the levee. Sandbags are placed around the leak, and the water held by them rises until it is the same level as the river, at which point the water there stops rising.

Exercise:

Problem:

Why is it difficult to swim under water in the Great Salt Lake?

Exercise:

Problem:

Is there a net force on a dam due to atmospheric pressure? Explain your answer.

Exercise:

Problem:

Does atmospheric pressure add to the gas pressure in a rigid tank? In a toy balloon? When, in general, does atmospheric pressure *not* affect the total pressure in a fluid?

Exercise:

Problem:

You can break a strong wine bottle by pounding a cork into it with your fist, but the cork must press directly against the liquid filling the bottle—there can be no air between the cork and liquid. Explain why the bottle breaks, and why it will not if there is air between the cork and liquid.

Problems & Exercises

Exercise:

Problem: What depth of mercury creates a pressure of 1.00 atm?

Solution:

0.760 m

Exercise:

Problem:

The greatest ocean depths on the Earth are found in the Marianas Trench near the Philippines. Calculate the pressure due to the ocean at the bottom of this trench, given its depth is 11.0 km and assuming the density of seawater is constant all the way down.

Exercise:

Problem: Verify that the SI unit of $h\rho g$ is N/m^2.

Solution:
Equation:

$$\begin{aligned}(h\rho g)_{\text{units}} &= (\text{m})\left(\text{kg/m}^3\right)\left(\text{m/s}^2\right) = (\text{kg}\cdot\text{m}^2)/(\text{m}^3\cdot\text{s}^2)\\ &= \left(\text{kg}\cdot\text{m/s}^2\right)\left(1/\text{m}^2\right)\\ &= \text{N/m}^2\end{aligned}$$

Exercise:

Problem:

Water towers store water above the level of consumers for times of heavy use, eliminating the need for high-speed pumps. How high above a user must the water level be to create a gauge pressure of $3.00 \times 10^5\ \text{N/m}^2$?

Exercise:

Problem:

The aqueous humor in a person’s eye is exerting a force of 0.300 N on the 1.10-cm^2 area of the cornea. (a) What pressure is this in mm Hg? (b) Is this value within the normal range for pressures in the eye?

Solution:

(a) 20.5 mm Hg

(b) The range of pressures in the eye is 12–24 mm Hg, so the result in part (a) is within that range

Exercise:

Problem:

How much force is exerted on one side of an 8.50 cm by 11.0 cm sheet of paper by the atmosphere? How can the paper withstand such a force?

Exercise:

Problem:

What pressure is exerted on the bottom of a 0.500-m-wide by 0.900-m-long gas tank that can hold 50.0 kg of gasoline by the weight of the gasoline in it when it is full?

Solution:

$1.09 \times 10^3\ \text{N/m}^2$

Exercise:

Problem:

Calculate the average pressure exerted on the palm of a shot-putter’s hand by the shot if the area of contact is $50.0\ \text{cm}^2$ and he exerts a force of 800 N on it. Express the pressure in N/m^2 and compare it with the $1.00 \times 10^6\ \text{Pa}$ pressures sometimes encountered in the skeletal system.

Exercise:

Problem:

The left side of the heart creates a pressure of 120 mm Hg by exerting a force directly on the blood over an effective area of $15.0\ \text{cm}^2$. What force does it exert to accomplish this?

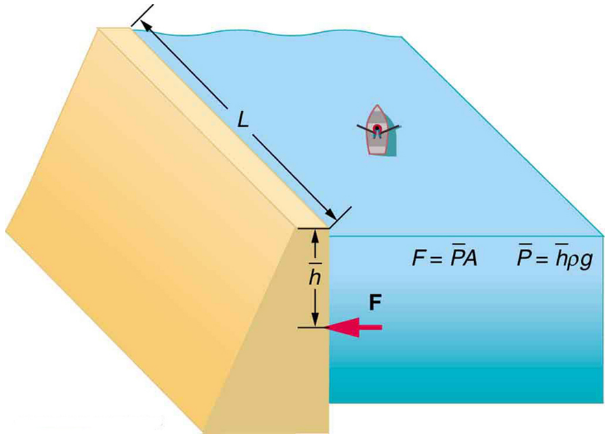
Solution:

24.0 N

Exercise:

Problem:

Show that the total force on a rectangular dam due to the water behind it increases with the *square* of the water depth. In particular, show that this force is given by $F = \rho g h^2 L/2$, where ρ is the density of water, h is its depth at the dam, and L is the length of the dam. You may assume the face of the dam is vertical. (Hint: Calculate the average pressure exerted and multiply this by the area in contact with the water. (See [link].)


L
$\bar{h}$
$\mathbf{F}$
$F = \bar{P}A$
$\bar{P} = \bar{h}\rho g$

Glossary

pressure

the weight of the fluid divided by the area supporting it

Pascal’s Principle

- Define pressure.
- State Pascal’s principle.
- Understand applications of Pascal’s principle.
- Derive relationships between forces in a hydraulic system.

Pressure is defined as force per unit area. Can pressure be increased in a fluid by pushing directly on the fluid? Yes, but it is much easier if the fluid is enclosed. The heart, for example, increases blood pressure by pushing directly on the blood in an enclosed system (valves closed in a chamber). If you try to push on a fluid in an open system, such as a river, the fluid flows away. An enclosed fluid cannot flow away, and so pressure is more easily increased by an applied force.

What happens to a pressure in an enclosed fluid? Since atoms in a fluid are free to move about, they transmit the pressure to all parts of the fluid and to the walls of the container. Remarkably, the pressure is transmitted *undiminished*. This phenomenon is called **Pascal’s principle**, because it was first clearly stated by the French philosopher and scientist Blaise Pascal (1623–1662): A change in pressure applied to an enclosed fluid is transmitted undiminished to all portions of the fluid and to the walls of its container.

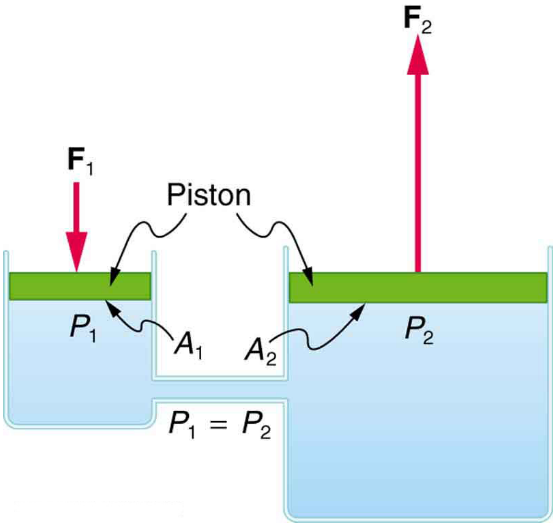
Note:
Pascal’s Principle
A change in pressure applied to an enclosed fluid is transmitted undiminished to all portions of the fluid and to the walls of its container.

Pascal’s principle, an experimentally verified fact, is what makes pressure so important in fluids. Since a change in pressure is transmitted undiminished in an enclosed fluid, we often know more about pressure than other physical quantities in fluids. Moreover, Pascal’s principle implies that *the total pressure in a fluid is the sum of the pressures from different sources*. We shall find this fact—that pressures add—very useful.

Blaise Pascal had an interesting life in that he was home-schooled by his father who removed all of the mathematics textbooks from his house and forbade him to study mathematics until the age of 15. This, of course, raised the boy's curiosity, and by the age of 12, he started to teach himself geometry. Despite this early deprivation, Pascal went on to make major contributions in the mathematical fields of probability theory, number theory, and geometry. He is also well known for being the inventor of the first mechanical digital calculator, in addition to his contributions in the field of fluid statics.

Application of Pascal's Principle

One of the most important technological applications of Pascal's principle is found in a *hydraulic system*, which is an enclosed fluid system used to exert forces. The most common hydraulic systems are those that operate car brakes. Let us first consider the simple hydraulic system shown in [link].



A typical hydraulic system with two fluid-filled cylinders, capped with pistons and connected by a tube called a hydraulic line. A downward force $\mathbf{F}_1$ on the left piston creates a pressure that is transmitted undiminished to all parts of the enclosed fluid. This results in an upward force $\mathbf{F}_2$ on the right piston that is larger than $\mathbf{F}_1$ because the right piston has a larger area.

Relationship Between Forces in a Hydraulic System

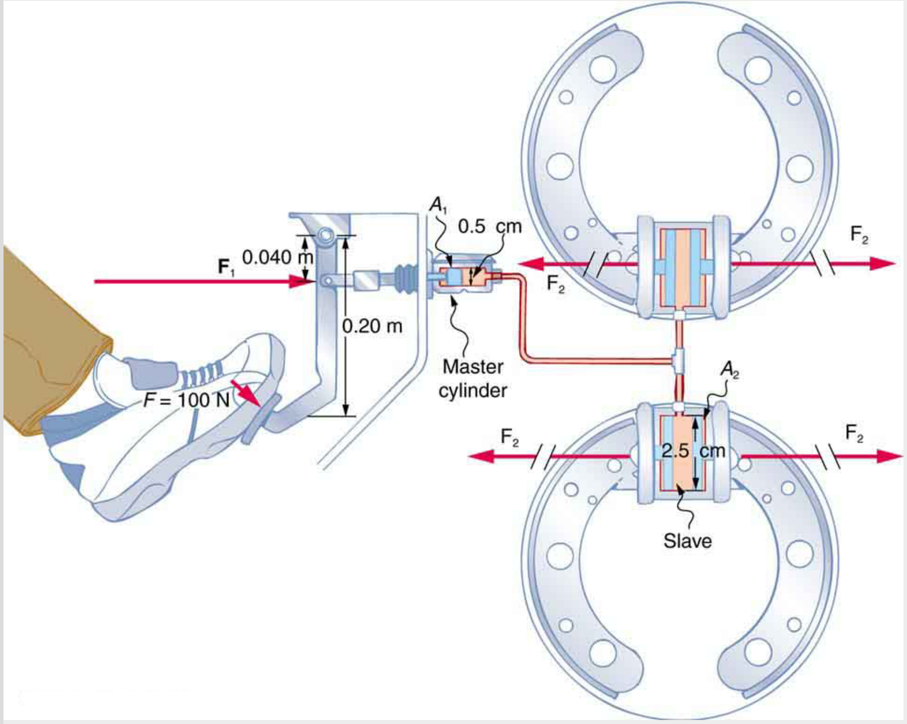
We can derive a relationship between the forces in the simple hydraulic system shown in [link] by applying Pascal's principle. Note first that the two pistons in the system are at the same height, and so there will be no difference in pressure due to a difference in depth. Now the pressure due to F_1 acting on area A_1 is simply $P_1 = \frac{F_1}{A_1}$, as defined by $P = \frac{F}{A}$. According to Pascal's principle, this pressure is transmitted undiminished throughout the fluid and to all walls of the container. Thus, a pressure P_2 is felt at the other piston that is equal to P_1. That is $P_1 = P_2$.

But since $P_2 = \frac{F_2}{A_2}$, we see that $\frac{F_1}{A_1} = \frac{F_2}{A_2}$.

This equation relates the ratios of force to area in any hydraulic system, providing the pistons are at the same vertical height and that friction in the system is negligible. Hydraulic systems can increase or decrease the force applied to them. To make the force larger, the pressure is applied to a larger area. For example, if a 100-N force is applied to the left cylinder in [link] and the right one has an area five times greater, then the force out is 500 N. Hydraulic systems are analogous to simple levers, but they have the advantage that pressure can be sent through tortuously curved lines to several places at once.

Example:
Calculating Force of Slave Cylinders: Pascal Puts on the Brakes

Consider the automobile hydraulic system shown in [link].



Hydraulic brakes use Pascal’s principle. The driver exerts a force of 100 N on the brake pedal. This force is increased by the simple lever and again by the hydraulic system. Each of the identical slave cylinders receives the same pressure and, therefore, creates the same force output F_2. The circular cross-sectional areas of the master and slave cylinders are represented by A_1 and A_2, respectively

A force of 100 N is applied to the brake pedal, which acts on the cylinder—called the master—through a lever. A force of 500 N is exerted on the master cylinder. (The reader can verify that the force is 500 N using techniques of statics from Applications of Statics, Including Problem-Solving Strategies.) Pressure created in the master cylinder is transmitted to four so-called slave cylinders. The master cylinder has a diameter of

0.500 cm, and each slave cylinder has a diameter of 2.50 cm. Calculate the force F_2 created at each of the slave cylinders.

Strategy

We are given the force F_1 that is applied to the master cylinder. The cross-sectional areas A_1 and A_2 can be calculated from their given diameters. Then $\frac{F_1}{A_1} = \frac{F_2}{A_2}$ can be used to find the force F_2. Manipulate this algebraically to get F_2 on one side and substitute known values:

Solution

Pascal's principle applied to hydraulic systems is given by $\frac{F_1}{A_1} = \frac{F_2}{A_2}$:

Equation:

$$F_2 = \frac{A_2}{A_1} F_1 = \frac{\pi r_2^2}{\pi r_1^2} F_1 = \frac{(1.25 \text{ cm})^2}{(0.250 \text{ cm})^2} \times 500 \text{ N} = 1.25 \times 10^4 \text{ N}.$$

Discussion

This value is the force exerted by each of the four slave cylinders. Note that we can add as many slave cylinders as we wish. If each has a 2.50-cm diameter, each will exert 1.25×10^4 N.

A simple hydraulic system, such as a simple machine, can increase force but cannot do more work than done on it. Work is force times distance moved, and the slave cylinder moves through a smaller distance than the master cylinder. Furthermore, the more slaves added, the smaller the distance each moves. Many hydraulic systems—such as power brakes and those in bulldozers—have a motorized pump that actually does most of the work in the system. The movement of the legs of a spider is achieved partly by hydraulics. Using hydraulics, a jumping spider can create a force that makes it capable of jumping 25 times its length!

Note:

Making Connections: Conservation of Energy

Conservation of energy applied to a hydraulic system tells us that the system cannot do more work than is done on it. Work transfers energy, and so the work output cannot exceed the work input. Power brakes and other similar hydraulic systems use pumps to supply extra energy when needed.

Section Summary

- Pressure is force per unit area.
- A change in pressure applied to an enclosed fluid is transmitted undiminished to all portions of the fluid and to the walls of its container.
- A hydraulic system is an enclosed fluid system used to exert forces.

Conceptual Questions

Exercise:

Problem:

Suppose the master cylinder in a hydraulic system is at a greater height than the slave cylinder. Explain how this will affect the force produced at the slave cylinder.

Problems & Exercises

Exercise:

Problem:

How much pressure is transmitted in the hydraulic system considered in [link]? Express your answer in pascals and in atmospheres.

Solution:

2.55×10^7 Pa; or 251 atm

Exercise:

Problem:

What force must be exerted on the master cylinder of a hydraulic lift to support the weight of a 2000-kg car (a large car) resting on the slave cylinder? The master cylinder has a 2.00-cm diameter and the slave has a 24.0-cm diameter.

Exercise:

Problem:

A crass host pours the remnants of several bottles of wine into a jug after a party. He then inserts a cork with a 2.00-cm diameter into the bottle, placing it in direct contact with the wine. He is amazed when he pounds the cork into place and the bottom of the jug (with a 14.0-cm diameter) breaks away. Calculate the extra force exerted against the bottom if he pounded the cork with a 120-N force.

Solution:

5.76×10^3 N extra force

Exercise:

Problem:

A certain hydraulic system is designed to exert a force 100 times as large as the one put into it. (a) What must be the ratio of the area of the slave cylinder to the area of the master cylinder? (b) What must be the ratio of their diameters? (c) By what factor is the distance through which the output force moves reduced relative to the distance through which the input force moves? Assume no losses to friction.

Exercise:

Problem:

(a) Verify that work input equals work output for a hydraulic system assuming no losses to friction. Do this by showing that the distance the output force moves is reduced by the same factor that the output force is increased. Assume the volume of the fluid is constant. (b) What effect would friction within the fluid and between components in the system have on the output force? How would this depend on whether or not the fluid is moving?

Solution:

(a) $V = d_\mathrm{i} A_\mathrm{i} = d_\mathrm{o} A_\mathrm{o} \Rightarrow d_\mathrm{o} = d_\mathrm{i}\left(\frac{A_\mathrm{i}}{A_\mathrm{o}}\right)$.

Now, using equation:
Equation:

$$\frac{F_1}{A_1} = \frac{F_2}{A_2} \Rightarrow F_\mathrm{o} = F_\mathrm{i}\left(\frac{A_\mathrm{o}}{A_\mathrm{i}}\right).$$

Finally,
Equation:

$$W_\mathrm{o} = F_\mathrm{o} d_\mathrm{o} = \left(\frac{F_\mathrm{i} A_\mathrm{o}}{A_\mathrm{i}}\right)\left(\frac{d_\mathrm{i} A_\mathrm{i}}{A_\mathrm{o}}\right) = F_\mathrm{i} d_\mathrm{i} = W_\mathrm{i}.$$

In other words, the work output equals the work input.

(b) If the system is not moving, friction would not play a role. With friction, we know there are losses, so that $W_\mathrm{out} = W_\mathrm{in} - W_\mathrm{f}$; therefore, the work output is less than the work input. In other words, with friction, you need to push harder on the input piston than was calculated for the nonfriction case.

Glossary

Pascal’s Principle

a change in pressure applied to an enclosed fluid is transmitted undiminished to all portions of the fluid and to the walls of its container

Gauge Pressure, Absolute Pressure, and Pressure Measurement

- Define gauge pressure and absolute pressure.
- Understand the working of aneroid and open-tube barometers.

If you limp into a gas station with a nearly flat tire, you will notice the tire gauge on the airline reads nearly zero when you begin to fill it. In fact, if there were a gaping hole in your tire, the gauge would read zero, even though atmospheric pressure exists in the tire. Why does the gauge read zero? There is no mystery here. Tire gauges are simply designed to read zero at atmospheric pressure and positive when pressure is greater than atmospheric.

Similarly, atmospheric pressure adds to blood pressure in every part of the circulatory system. (As noted in Pascal’s Principle, the total pressure in a fluid is the sum of the pressures from different sources—here, the heart and the atmosphere.) But atmospheric pressure has no net effect on blood flow since it adds to the pressure coming out of the heart and going back into it, too. What is important is how much *greater* blood pressure is than atmospheric pressure. Blood pressure measurements, like tire pressures, are thus made relative to atmospheric pressure.

In brief, it is very common for pressure gauges to ignore atmospheric pressure—that is, to read zero at atmospheric pressure. We therefore define **gauge pressure** to be the pressure relative to atmospheric pressure. Gauge pressure is positive for pressures above atmospheric pressure, and negative for pressures below it.

Note:
Gauge Pressure
Gauge pressure is the pressure relative to atmospheric pressure. Gauge pressure is positive for pressures above atmospheric pressure, and negative for pressures below it.

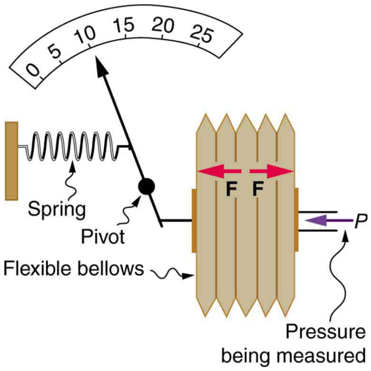
In fact, atmospheric pressure does add to the pressure in any fluid not enclosed in a rigid container. This happens because of Pascal’s principle. The total pressure, or **absolute pressure**, is thus the sum of gauge pressure and atmospheric pressure: $P_{\text{abs}} = P_{\text{g}} + P_{\text{atm}}$ where P_{abs} is absolute pressure, P_{g} is gauge pressure, and P_{atm} is atmospheric pressure. For example, if your tire gauge reads 34 psi (pounds per square inch), then the absolute pressure is 34 psi plus 14.7 psi (P_{atm} in psi), or 48.7 psi (equivalent to 336 kPa).

Note:
Absolute Pressure
Absolute pressure is the sum of gauge pressure and atmospheric pressure.

For reasons we will explore later, in most cases the absolute pressure in fluids cannot be negative. Fluids push rather than pull, so the smallest absolute pressure is zero. (A negative absolute pressure is a pull.) Thus the smallest possible gauge pressure is $P_{\text{g}} = -P_{\text{atm}}$ (this makes P_{abs} zero). There is no theoretical limit to how large a gauge pressure can be.

There are a host of devices for measuring pressure, ranging from tire gauges to blood pressure cuffs. Pascal's principle is of major importance in these devices. The undiminished transmission of pressure through a fluid allows precise remote sensing of pressures. Remote sensing is often more convenient than putting a measuring device into a system, such as a person's artery.

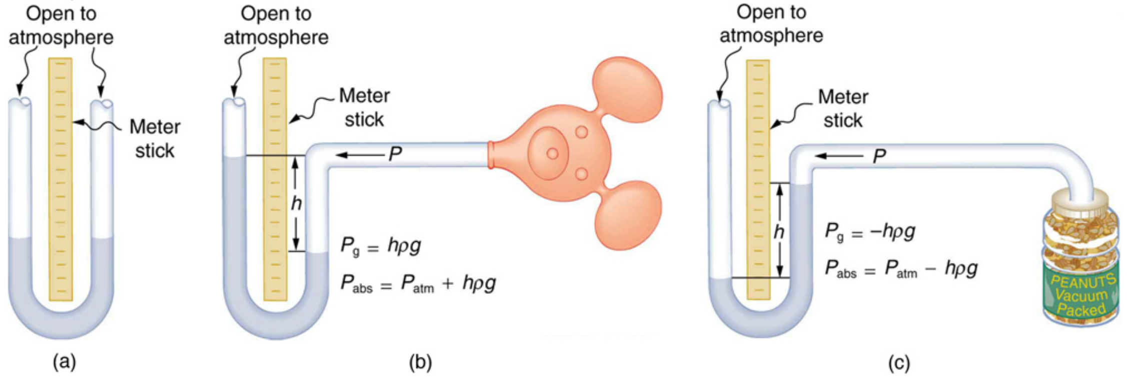
[link] shows one of the many types of mechanical pressure gauges in use today. In all mechanical pressure gauges, pressure results in a force that is converted (or transduced) into some type of readout.



This aneroid gauge utilizes flexible bellows connected to a mechanical indicator to measure pressure.

An entire class of gauges uses the property that pressure due to the weight of a fluid is given by $P = h\rho g$. Consider the U-shaped tube shown in [link], for example. This simple tube is called a *manometer*. In [link](a), both sides of the tube are open to the atmosphere. Atmospheric pressure therefore pushes down on each side equally so its effect cancels. If the fluid is deeper on one side, there is a greater pressure on the deeper side, and the fluid flows away from that side until the depths are equal.

Let us examine how a manometer is used to measure pressure. Suppose one side of the U-tube is connected to some source of pressure P_{abs} such as the toy balloon in [link](b) or the vacuum-packed peanut jar shown in [link](c). Pressure is transmitted undiminished to the manometer, and the fluid levels are no longer equal. In [link](b), P_{abs} is greater than atmospheric pressure, whereas in [link](c), P_{abs} is less than atmospheric pressure. In both cases, P_{abs} differs from atmospheric pressure by an amount $h\rho g$, where ρ is the density of the fluid in the manometer. In [link](b), P_{abs} can support a column of fluid of height h, and so it must exert a pressure $h\rho g$ greater than atmospheric pressure (the gauge pressure P_{g} is positive). In [link](c), atmospheric pressure can support a column of fluid of height h, and so P_{abs} is less than atmospheric pressure by an amount $h\rho g$ (the gauge pressure P_{g} is negative). A manometer with one side open to the atmosphere is an ideal device for measuring gauge pressures. The gauge pressure is $P_{\text{g}} = h\rho g$ and is found by measuring h.



An open-tube manometer has one side open to the atmosphere. (a) Fluid depth must be the same on both sides, or the pressure each side exerts at the bottom will be unequal and there will be flow from the deeper side. (b) A positive gauge pressure $P_{\text{g}} = h\rho g$ transmitted to one side of the manometer can support a column of fluid of height h. (c) Similarly, atmospheric pressure is greater than a negative gauge pressure P_{g} by an amount $h\rho g$. The jar's rigidity prevents atmospheric pressure from being transmitted to the peanuts.

Mercury manometers are often used to measure arterial blood pressure. An inflatable cuff is placed on the upper arm as shown in [link]. By squeezing the bulb, the person making the measurement exerts pressure, which is transmitted undiminished to both the main artery in the arm and the manometer. When this applied pressure exceeds blood pressure, blood flow below the cuff is cut off. The person making the measurement then slowly lowers the applied pressure and listens for blood flow to resume. Blood pressure pulsates because of the pumping action of the heart, reaching a maximum, called **systolic pressure**, and a minimum, called **diastolic pressure**, with each heartbeat. Systolic pressure is measured by noting the value of h when blood flow first begins as cuff pressure is lowered. Diastolic pressure is measured by noting h when blood flows without interruption. The typical blood pressure of a young adult raises the mercury to a height of 120 mm at systolic and 80 mm at diastolic. This is commonly quoted as 120 over 80, or 120/80. The first pressure is representative of the maximum output of the heart; the second is due to the elasticity of the arteries in maintaining the pressure between beats. The density of the mercury fluid in the manometer is 13.6 times greater than water, so the height of the fluid will be 1/13.6 of that in a water manometer. This reduced height can make measurements difficult, so mercury manometers are used to measure larger pressures, such as blood pressure. The density of mercury is such that $1.0 \text{ mm Hg} = 133 \text{ Pa}$.

Note:
Systolic Pressure
Systolic pressure is the maximum blood pressure.

Note:
Diastolic Pressure
Diastolic pressure is the minimum blood pressure.

In routine blood pressure measurements, an inflatable cuff is placed on the upper arm at the same level as the heart. Blood flow is detected just below the cuff, and corresponding pressures are transmitted to a mercury-filled manometer. (credit: U.S. Army photo by Spc. Micah E. Clare\4TH BCT)

Example:
Calculating Height of IV Bag: Blood Pressure and Intravenous Infusions

Intravenous infusions are usually made with the help of the gravitational force. Assuming that the density of the fluid being administered is 1.00 g/ml, at what height should the IV bag be placed above the entry point so that the fluid just enters the vein if the blood pressure in the vein is 18 mm Hg above atmospheric pressure? Assume that the IV bag is collapsible.

Strategy for (a)

For the fluid to just enter the vein, its pressure at entry must exceed the blood pressure in the vein (18 mm Hg above atmospheric pressure). We therefore need to find the height of fluid that corresponds to this gauge pressure.

Solution

We first need to convert the pressure into SI units. Since $1.0\text{ mm Hg}=133\text{ Pa}$,

Equation:

$$P=18\text{ mm Hg}\times\frac{133\text{ Pa}}{1.0\text{ mm Hg}}=2400\text{ Pa}.$$

Rearranging $P_{\text{g}}=h\rho g$ for h gives $h=\frac{P_{\text{g}}}{\rho g}$. Substituting known values into this equation gives

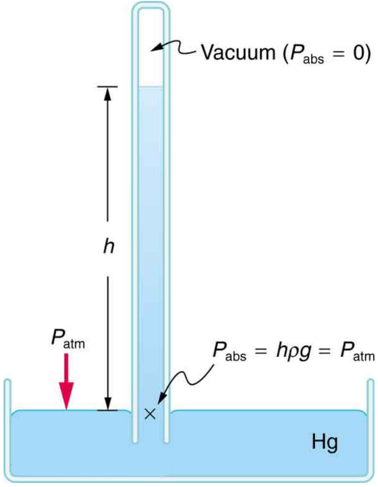
Equation:

$$\begin{aligned} h &= \frac{2400\text{ N/m}^2}{\left(1.0\times10^3\text{ kg/m}^3\right)\left(9.80\text{ m/s}^2\right)} \\ &= 0.24\text{ m}. \end{aligned}$$

Discussion

The IV bag must be placed at 0.24 m above the entry point into the arm for the fluid to just enter the arm. Generally, IV bags are placed higher than this. You may have noticed that the bags used for blood collection are placed below the donor to allow blood to flow easily from the arm to the bag, which is the opposite direction of flow than required in the example presented here.

A *barometer* is a device that measures atmospheric pressure. A mercury barometer is shown in [link]. This device measures atmospheric pressure, rather than gauge pressure, because there is a nearly pure vacuum above the mercury in the tube. The height of the mercury is such that $h\rho g=P_{\text{atm}}$. When atmospheric pressure varies, the mercury rises or falls, giving important clues to weather forecasters. The barometer can also be used as an altimeter, since average atmospheric pressure varies with altitude. Mercury barometers and manometers are so common that units of mm Hg are often quoted for atmospheric pressure and blood pressures. [link] gives conversion factors for some of the more commonly used units of pressure.



A mercury barometer measures atmospheric pressure. The pressure due to the mercury's weight, $h\rho g$, equals atmospheric pressure. The atmosphere is able to force mercury in the tube to a height h because the pressure above the mercury is zero.

Conversion to N/m^2 (Pa)	Conversion from atm
$1.0\text{ atm} = 1.013 \times 10^5\text{ N/m}^2$	$1.0\text{ atm} = 1.013 \times 10^5\text{ N/m}^2$
$1.0\text{ dyne/cm}^2 = 0.10\text{ N/m}^2$	$1.0\text{ atm} = 1.013 \times 10^6\text{ dyne/cm}^2$
$1.0\text{ kg/cm}^2 = 9.8 \times 10^4\text{ N/m}^2$	$1.0\text{ atm} = 1.013\text{ kg/cm}^2$
$1.0\text{ lb/in.}^2 = 6.90 \times 10^3\text{ N/m}^2$	$1.0\text{ atm} = 14.7\text{ lb/in.}^2$
$1.0\text{ mm Hg} = 133\text{ N/m}^2$	$1.0\text{ atm} = 760\text{ mm Hg}$
$1.0\text{ cm Hg} = 1.33 \times 10^3\text{ N/m}^2$	$1.0\text{ atm} = 76.0\text{ cm Hg}$
$1.0\text{ cm water} = 98.1\text{ N/m}^2$	$1.0\text{ atm} = 1.03 \times 10^3\text{ cm water}$
$1.0\text{ bar} = 1.000 \times 10^5\text{ N/m}^2$	$1.0\text{ atm} = 1.013\text{ bar}$
$1.0\text{ millibar} = 1.000 \times 10^2\text{ N/m}^2$	$1.0\text{ atm} = 1013\text{ millibar}$

Conversion Factors for Various Pressure Units

Section Summary

- Gauge pressure is the pressure relative to atmospheric pressure.
- Absolute pressure is the sum of gauge pressure and atmospheric pressure.
- Aneroid gauge measures pressure using a bellows-and-spring arrangement connected to the pointer of a calibrated scale.
- Open-tube manometers have U-shaped tubes and one end is always open. It is used to measure pressure.
- A mercury barometer is a device that measures atmospheric pressure.

Conceptual Questions

Exercise:

Problem:

Explain why the fluid reaches equal levels on either side of a manometer if both sides are open to the atmosphere, even if the tubes are of different diameters.

Exercise:

Problem:

[link] shows how a common measurement of arterial blood pressure is made. Is there any effect on the measured pressure if the manometer is lowered? What is the effect of raising the arm above the shoulder? What is the effect of placing the cuff on the upper leg with the person standing? Explain your answers in terms of pressure created by the weight of a fluid.

Exercise:

Problem:

Considering the magnitude of typical arterial blood pressures, why are mercury rather than water manometers used for these measurements?

Problems & Exercises

Exercise:

Problem:

Find the gauge and absolute pressures in the balloon and peanut jar shown in [link], assuming the manometer connected to the balloon uses water whereas the manometer connected to the jar contains mercury. Express in units of centimeters of water for the balloon and millimeters of mercury for the jar, taking $h = 0.0500\text{ m}$ for each.

Solution:

Balloon:

$$\begin{aligned} P_{\text{g}} &= 5.00\text{ cm H}_2\text{O}, \\ P_{\text{abs}} &= 1.035 \times 10^3\text{ cm H}_2\text{O}. \end{aligned}$$

Jar:

$$\begin{aligned} P_{\text{g}} &= -50.0\text{ mm Hg}, \\ P_{\text{abs}} &= 710\text{ mm Hg}. \end{aligned}$$

Exercise:

Problem:

(a) Convert normal blood pressure readings of 120 over 80 mm Hg to newtons per meter squared using the relationship for pressure due to the weight of a fluid $(P = h\rho g)$ rather than a conversion factor. (b) Discuss why blood pressures for an infant could be smaller than those for an adult. Specifically, consider the smaller height to which blood must be pumped.

Exercise:

Problem:

How tall must a water-filled manometer be to measure blood pressures as high as 300 mm Hg?

Solution:

4.08 m

Exercise:

Problem:

Pressure cookers have been around for more than 300 years, although their use has strongly declined in recent years (early models had a nasty habit of exploding). How much force must the latches holding the lid onto a pressure cooker be able to withstand if the circular lid is 25.0 cm in diameter and the gauge pressure inside is 300 atm? Neglect the weight of the lid.

Exercise:

Problem:

Suppose you measure a standing person’s blood pressure by placing the cuff on his leg 0.500 m below the heart. Calculate the pressure you would observe (in units of mm Hg) if the pressure at the heart were 120 over 80 mm Hg. Assume that there is no loss of pressure due to resistance in the circulatory system (a reasonable assumption, since major arteries are large).

Solution:

$$\Delta P = 38.7\text{ mm Hg},$$

$$\text{Leg blood pressure} = \frac{159}{119}.$$

Exercise:

Problem:

A submarine is stranded on the bottom of the ocean with its hatch 25.0 m below the surface. Calculate the force needed to open the hatch from the inside, given it is circular and 0.450 m in diameter. Air pressure inside the submarine is 1.00 atm.

Exercise:

Problem:

Assuming bicycle tires are perfectly flexible and support the weight of bicycle and rider by pressure alone, calculate the total area of the tires in contact with the ground. The bicycle plus rider has a mass of 80.0 kg, and the gauge pressure in the tires is $3.50 \times 10^5\text{ Pa}$.

Solution:

$22.4\ \text{cm}^2$

Glossary

absolute pressure
the sum of gauge pressure and atmospheric pressure

diastolic pressure
the minimum blood pressure in the artery

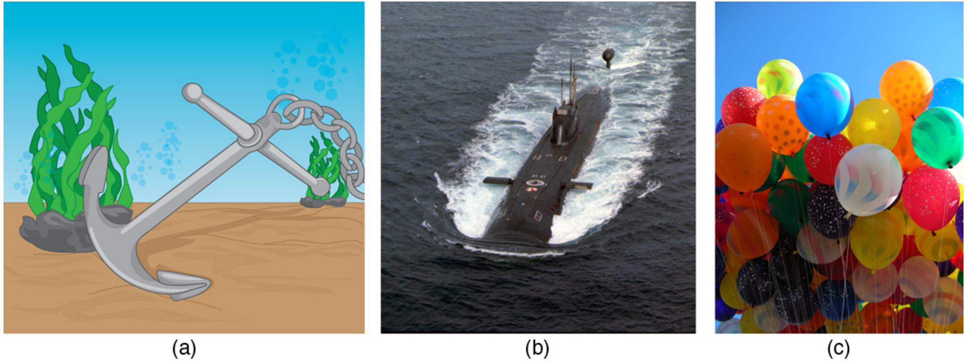
gauge pressure
the pressure relative to atmospheric pressure

systolic pressure
the maximum blood pressure in the artery

Archimedes’ Principle

- Define buoyant force.
- State Archimedes’ principle.
- Understand why objects float or sink.
- Understand the relationship between density and Archimedes’ principle.

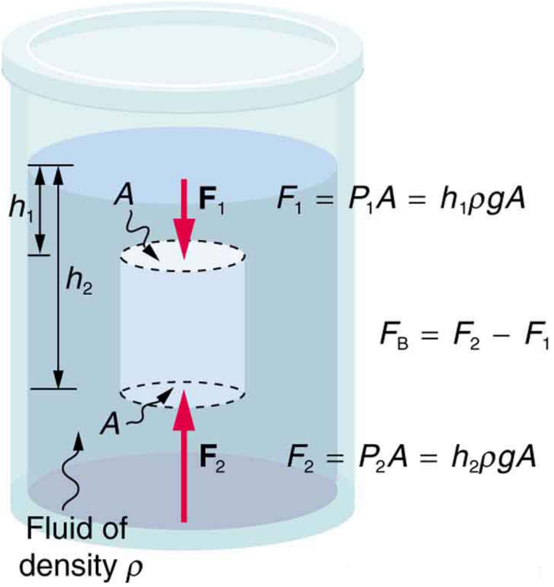
When you rise from lounging in a warm bath, your arms feel strangely heavy. This is because you no longer have the buoyant support of the water. Where does this buoyant force come from? Why is it that some things float and others do not? Do objects that sink get any support at all from the fluid? Is your body buoyed by the atmosphere, or are only helium balloons affected? (See [link].)



(a) Even objects that sink, like this anchor, are partly supported by water when submerged. (b) Submarines have adjustable density (ballast tanks) so that they may float or sink as desired. (credit: Allied Navy) (c) Helium-filled balloons tug upward on their strings, demonstrating air’s buoyant effect. (credit: Crystl)

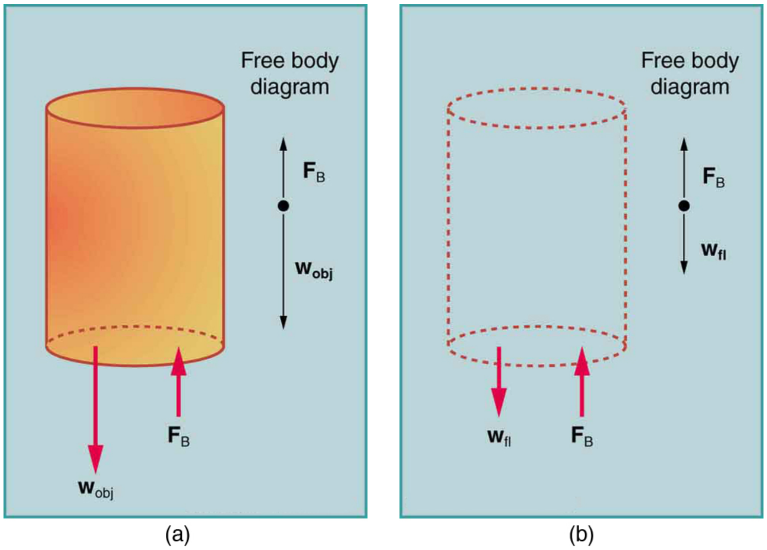
Answers to all these questions, and many others, are based on the fact that pressure increases with depth in a fluid. This means that the upward force on the bottom of an object in a fluid is greater than the downward force on the top of the object. There is a net upward, or **buoyant force** on any object in any fluid. (See [link].) If the buoyant force is greater than the object’s weight, the object will rise to the surface and float. If the buoyant force is less than the object's weight, the object will sink. If the buoyant force equals the object's weight, the object will remain suspended at that depth. The buoyant force is always present whether the object floats, sinks, or is suspended in a fluid.

Note:
Buoyant Force
The buoyant force is the net upward force on any object in any fluid.


h_1
h_2
A
$\mathbf{F}_1$
$F_1 = P_1A = h_1\rho gA$
$F_B = F_2 - F_1$
A
$\mathbf{F}_2$
$F_2 = P_2A = h_2\rho gA$
Fluid of density ρ

Pressure due to the weight of a fluid increases with depth since $P = h\rho g$. This pressure and associated upward force on the bottom of the cylinder are greater than the downward force on the top of the cylinder. Their difference is the buoyant force $\mathbf{F}_\mathrm{B}$. (Horizontal forces cancel.)

Just how great is this buoyant force? To answer this question, think about what happens when a submerged object is removed from a fluid, as in [link].



(a) An object submerged in a fluid experiences a buoyant force F_B. If F_B is greater than the weight of the object, the object will rise. If F_B is less than the weight of the object, the object will sink. (b) If the object is removed, it is replaced by fluid having weight w_fl. Since this weight is supported by surrounding fluid, the buoyant force must equal the weight of the fluid displaced. That is, $F_\mathrm{B} = w_\mathrm{fl}$,a statement of Archimedes' principle.

The space it occupied is filled by fluid having a weight w_fl. This weight is supported by the surrounding fluid, and so the buoyant force must equal w_fl, the weight of the fluid displaced by the object. It is a tribute to the genius of the Greek mathematician and inventor Archimedes (ca. 287–212 B.C.) that he stated this principle long before concepts of force were well established. Stated in words, **Archimedes' principle** is as follows: The buoyant force on an object equals the weight of the fluid it displaces. In equation form, Archimedes' principle is

Equation:

$$F_{\mathrm{B}} = w_{\mathrm{fl}},$$

where F_{B} is the buoyant force and w_{fl} is the weight of the fluid displaced by the object. Archimedes' principle is valid in general, for any object in any fluid, whether partially or totally submerged.

Note:

Archimedes' Principle

According to this principle the buoyant force on an object equals the weight of the fluid it displaces. In equation form, Archimedes' principle is

Equation:

$$F_{\mathrm{B}} = w_{\mathrm{fl}},$$

where F_{B} is the buoyant force and w_{fl} is the weight of the fluid displaced by the object.

Humm … High-tech body swimsuits were introduced in 2008 in preparation for the Beijing Olympics. One concern (and international rule) was that these suits should not provide any buoyancy advantage. How do you think that this rule could be verified?

Note:

Making Connections: Take-Home Investigation

The density of aluminum foil is 2.7 times the density of water. Take a piece of foil, roll it up into a ball and drop it into water. Does it sink? Why or why not? Can you make it sink?

Floating and Sinking

Drop a lump of clay in water. It will sink. Then mold the lump of clay into the shape of a boat, and it will float. Because of its shape, the boat displaces more water than the lump and experiences a greater buoyant force. The same is true of steel ships.

Example:

Calculating buoyant force: dependency on shape

(a) Calculate the buoyant force on 10,000 metric tons $(1.00 \times 10^7 \text{ kg})$ of solid steel completely submerged in water, and compare this with the steel's weight. (b) What is the maximum buoyant force that water could exert on this same steel if it were shaped into a boat that could displace $1.00 \times 10^5 \text{ m}^3$ of water?

Strategy for (a)

To find the buoyant force, we must find the weight of water displaced. We can do this by using the densities of water and steel given in [link]. We note that, since the steel is completely submerged, its volume and the water's volume are the same. Once we know the volume of water, we can find its mass and weight.

Solution for (a)

First, we use the definition of density $\rho = \frac{m}{V}$ to find the steel's volume, and then we substitute values for mass and density. This gives

Equation:

$$V_{\text{st}} = \frac{m_{\text{st}}}{\rho_{\text{st}}} = \frac{1.00 \times 10^7 \text{ kg}}{7.8 \times 10^3 \text{ kg/m}^3} = 1.28 \times 10^3 \text{ m}^3.$$

Because the steel is completely submerged, this is also the volume of water displaced, V_{w}. We can now find the mass of water displaced from the relationship between its volume and density, both of which are known. This gives

Equation:

$$\begin{aligned} m_{\mathrm{w}} &= \rho_{\mathrm{w}} V_{\mathrm{w}} = (1.000 \times 10^3 \text{ kg/m}^3)(1.28 \times 10^3 \text{ m}^3) \\ &= 1.28 \times 10^6 \text{ kg.} \end{aligned}$$

By Archimedes’ principle, the weight of water displaced is $m_{\mathrm{w}}g$, so the buoyant force is

Equation:

$$\begin{aligned} F_{\mathrm{B}} &= w_{\mathrm{w}} = m_{\mathrm{w}} g = \left(1.28 \times 10^6 \text{ kg}\right)\left(9.80 \text{ m/s}^2\right) \\ &= 1.3 \times 10^7 \text{ N.} \end{aligned}$$

The steel’s weight is $m_{\mathrm{w}} g = 9.80 \times 10^7$ N, which is much greater than the buoyant force, so the steel will remain submerged. Note that the buoyant force is rounded to two digits because the density of steel is given to only two digits.

Strategy for (b)

Here we are given the maximum volume of water the steel boat can displace. The buoyant force is the weight of this volume of water.

Solution for (b)

The mass of water displaced is found from its relationship to density and volume, both of which are known. That is,

Equation:

$$\begin{aligned} m_{\mathrm{w}} &= \rho_{\mathrm{w}} V_{\mathrm{w}} = \left(1.000 \times 10^3 \text{ kg/m}^3\right)\left(1.00 \times 10^5 \text{ m}^3\right) \\ &= 1.00 \times 10^8 \text{ kg.} \end{aligned}$$

The maximum buoyant force is the weight of this much water, or

Equation:

$$\begin{aligned} F_\text{B} &= w_\text{w} = m_\text{w}g = \left(1.00 \times 10^8\text{ kg}\right)\left(9.80\text{ m/s}^2\right) \\ &= 9.80 \times 10^8\text{ N}. \end{aligned}$$

Discussion

The maximum buoyant force is ten times the weight of the steel, meaning the ship can carry a load nine times its own weight without sinking.

Note:

Making Connections: Take-Home Investigation

A piece of household aluminum foil is 0.016 mm thick. Use a piece of foil that measures 10 cm by 15 cm. (a) What is the mass of this amount of foil? (b) If the foil is folded to give it four sides, and paper clips or washers are added to this “boat,” what shape of the boat would allow it to hold the most “cargo” when placed in water? Test your prediction.

Density and Archimedes’ Principle

Density plays a crucial role in Archimedes’ principle. The average density of an object is what ultimately determines whether it floats. If its average density is less than that of the surrounding fluid, it will float. This is because the fluid, having a higher density, contains more mass and hence more weight in the same volume. The buoyant force, which equals the weight of the fluid displaced, is thus greater than the weight of the object. Likewise, an object denser than the fluid will sink.

The extent to which a floating object is submerged depends on how the object’s density is related to that of the fluid. In [link], for example, the unloaded ship has a lower density and less of it is submerged compared with the same ship loaded. We can derive a quantitative expression for the fraction submerged by considering density. The fraction submerged is the ratio of the volume submerged to the volume of the object, or

Equation:

$$\text{fraction submerged} = \frac{V_{\text{sub}}}{V_{\text{obj}}} = \frac{V_{\text{fl}}}{V_{\text{obj}}}.$$

The volume submerged equals the volume of fluid displaced, which we call V_{fl}. Now we can obtain the relationship between the densities by substituting $\rho = \frac{m}{V}$ into the expression. This gives

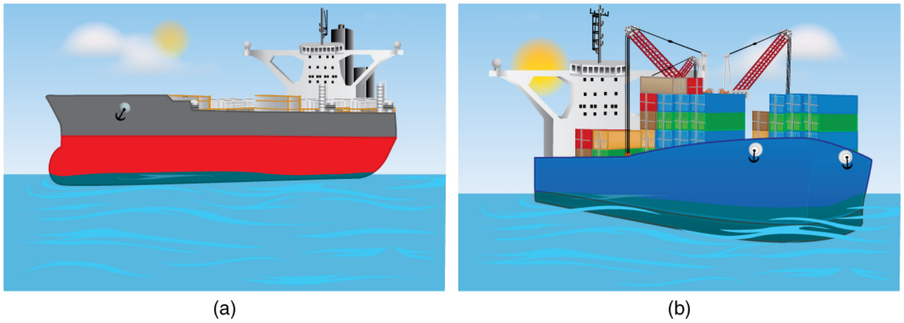
Equation:

$$\frac{V_{\text{fl}}}{V_{\text{obj}}} = \frac{m_{\text{fl}}/\rho_{\text{fl}}}{m_{\text{obj}}/\rho_{\text{obj}}},$$

where ρ_{obj} is the average density of the object and ρ_{fl} is the density of the fluid. Since the object floats, its mass and that of the displaced fluid are equal, and so they cancel from the equation, leaving

Equation:

$$\text{fraction submerged} = \frac{\rho_{\text{obj}}}{\rho_{\text{fl}}}.$$



An unloaded ship (a) floats higher in the water than a loaded ship (b).

We use this last relationship to measure densities. This is done by measuring the fraction of a floating object that is submerged—for example, with a hydrometer. It is useful to define the ratio of the density of an object to a fluid (usually water) as **specific gravity**:

Equation:

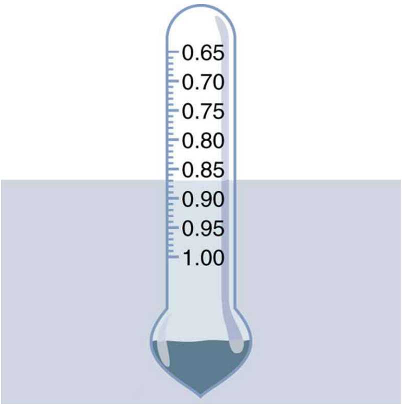
$$\text{specific gravity} = \frac{\rho}{\rho_{\text{w}}},$$

where ρ is the average density of the object or substance and ρ_{w} is the density of water at 4.00°C. Specific gravity is dimensionless, independent of whatever units are used for ρ. If an object floats, its specific gravity is less than one. If it sinks, its specific gravity is greater than one. Moreover, the fraction of a floating object that is submerged equals its specific gravity. If an object's specific gravity is exactly 1, then it will remain suspended in the fluid, neither sinking nor floating. Scuba divers try to obtain this state so that they can hover in the water. We measure the specific gravity of fluids, such as battery acid, radiator fluid, and urine, as an indicator of their condition. One device for measuring specific gravity is shown in [link].

Note:

Specific Gravity

Specific gravity is the ratio of the density of an object to a fluid (usually water).



This hydrometer is floating in a fluid of specific gravity 0.87. The glass hydrometer is filled with air and weighted with lead at the bottom. It floats highest in the densest fluids and has been calibrated and labeled so that specific gravity can be read from it directly.

Example:

Calculating Average Density: Floating Woman

Suppose a 60.0-kg woman floats in freshwater with 97.0% of her volume submerged when her lungs are full of air. What is her average density?

Strategy

We can find the woman’s density by solving the equation

Equation:

$$\text{fraction submerged} = \frac{\rho_{\text{obj}}}{\rho_{\text{fl}}}$$

for the density of the object. This yields

Equation:

$$\rho_{\text{obj}} = \rho_{\text{person}} = (\text{fraction submerged}) \cdot \rho_{\text{fl}}.$$

We know both the fraction submerged and the density of water, and so we can calculate the woman’s density.

Solution

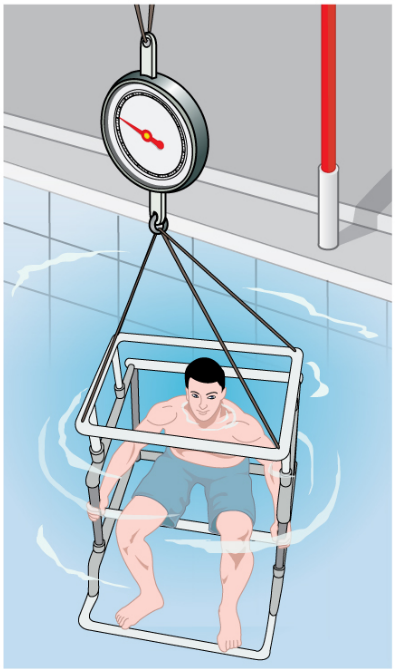
Entering the known values into the expression for her density, we obtain

Equation:

$$\rho_{\text{person}} = 0.970 \cdot \left(10^3 \frac{\text{kg}}{\text{m}^3}\right) = 970 \frac{\text{kg}}{\text{m}^3}.$$

Discussion

Her density is less than the fluid density. We expect this because she floats. Body density is one indicator of a person’s percent body fat, of interest in medical diagnostics and athletic training. (See [link].)

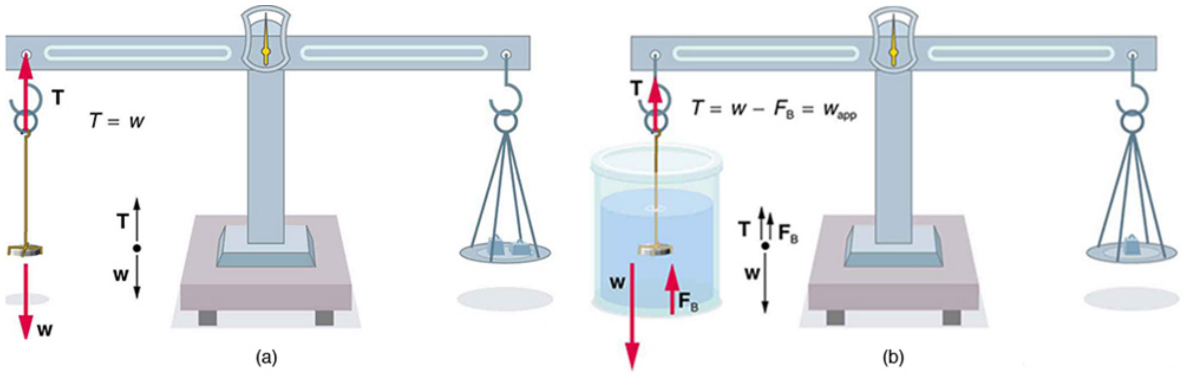


Subject in a “fat tank,” where he is weighed while completely submerged as part of a body density determination. The subject must completely empty his lungs and hold a metal weight in order to sink. Corrections are made for the residual air in his lungs (measured separately) and the metal weight. His corrected submerged weight, his weight in air, and pinch tests of strategic fatty areas are used to calculate his percent body fat.

There are many obvious examples of lower-density objects or substances floating in higher-density fluids—oil on water, a hot-air balloon, a bit of cork in wine, an iceberg, and hot wax in a “lava lamp,” to name a few. Less obvious examples include lava rising in a volcano and mountain ranges floating on the higher-density crust and mantle beneath them. Even seemingly solid Earth has fluid characteristics.

More Density Measurements

One of the most common techniques for determining density is shown in [link].



(a) A coin is weighed in air. (b) The apparent weight of the coin is determined while it is completely submerged in a fluid of known density. These two measurements are used to calculate the density of the coin.

An object, here a coin, is weighed in air and then weighed again while submerged in a liquid. The density of the coin, an indication of its authenticity, can be calculated if the fluid density is known. This same technique can also be used to determine the density of the fluid if the density of the coin is known. All of these calculations are based on Archimedes’ principle.

Archimedes’ principle states that the buoyant force on the object equals the weight of the fluid displaced. This, in turn, means that the object *appears* to weigh less when submerged; we call this measurement the object’s *apparent weight*. The object suffers an *apparent weight loss* equal to the weight of the fluid displaced. Alternatively, on balances that measure mass, the object suffers an *apparent mass loss* equal to the mass of fluid displaced. That is

Equation:

$$\text{apparent weight loss} = \text{weight of fluid displaced}$$

or

Equation:

$$\text{apparent mass loss} = \text{mass of fluid displaced.}$$

The next example illustrates the use of this technique.

Example:

Calculating Density: Is the Coin Authentic?

The mass of an ancient Greek coin is determined in air to be 8.630 g. When the coin is submerged in water as shown in [link], its apparent mass is 7.800 g. Calculate its density, given that water has a density of $1.000\ \text{g/cm}^3$ and that effects caused by the wire suspending the coin are negligible.

Strategy

To calculate the coin’s density, we need its mass (which is given) and its volume. The volume of the coin equals the volume of water displaced. The volume of water displaced V_{w} can be found by solving the equation for density $\rho = \frac{m}{V}$ for V.

Solution

The volume of water is $V_{\text{w}} = \frac{m_{\text{w}}}{\rho_{\text{w}}}$ where m_{w} is the mass of water displaced. As noted, the mass of the water displaced equals the apparent mass loss, which is $m_{\text{w}} = 8.630\text{ g} - 7.800\text{ g} = 0.830\text{ g}$. Thus the volume of water is $V_{\text{w}} = \frac{0.830\text{ g}}{1.000\text{ g/cm}^3} = 0.830\text{ cm}^3$. This is also the volume of the coin, since it is completely submerged. We can now find the density of the coin using the definition of density:

Equation:

$$\rho_{\text{c}} = \frac{m_{\text{c}}}{V_{\text{c}}} = \frac{8.630\text{ g}}{0.830\text{ cm}^3} = 10.4\text{ g/cm}^3.$$

Discussion

You can see from [link] that this density is very close to that of pure silver, appropriate for this type of ancient coin. Most modern counterfeits are not pure silver.

This brings us back to Archimedes’ principle and how it came into being. As the story goes, the king of Syracuse gave Archimedes the task of determining whether the royal crown maker was supplying a crown of pure gold. The purity of gold is difficult to determine by color (it can be diluted with other metals and still look as yellow as pure gold), and other analytical techniques had not yet been conceived. Even ancient peoples, however, realized that the density of gold was greater than that of any other then-known substance. Archimedes purportedly agonized over his task and had his inspiration one day while at the public baths, pondering the support the water gave his body. He came up with his now-famous principle, saw how to apply it to determine density, and ran naked down the streets of Syracuse crying “Eureka!” (Greek for “I have found it”). Similar behavior can be observed in contemporary physicists from time to time!

Note:

PhET Explorations: Buoyancy

When will objects float and when will they sink? Learn how buoyancy works with blocks. Arrows show the applied forces, and you can modify the properties of the blocks and the fluid.

https://phet.colorado.edu/sims/density-and-buoyancy/buoyancy_en.html

Section Summary

- Buoyant force is the net upward force on any object in any fluid. If the buoyant force is greater than the object’s weight, the object will rise to the surface and float. If the buoyant force is less than the object’s weight, the object will sink. If the buoyant force equals the object’s weight, the object will remain suspended at that depth. The buoyant force is always present whether the object floats, sinks, or is suspended in a fluid.
- Archimedes’ principle states that the buoyant force on an object equals the weight of the fluid it displaces.
- Specific gravity is the ratio of the density of an object to a fluid (usually water).

Conceptual Questions

Exercise:

Problem:

More force is required to pull the plug in a full bathtub than when it is empty. Does this contradict Archimedes’ principle? Explain your answer.

Exercise:

Problem:

Do fluids exert buoyant forces in a “weightless” environment, such as in the space shuttle? Explain your answer.

Exercise:

Problem:

Will the same ship float higher in salt water than in freshwater? Explain your answer.

Exercise:

Problem:

Marbles dropped into a partially filled bathtub sink to the bottom. Part of their weight is supported by buoyant force, yet the downward force on the bottom of the tub increases by exactly the weight of the marbles. Explain why.

Problem Exercises

Exercise:

Problem:

What fraction of ice is submerged when it floats in freshwater, given the density of water at 0°C is very close to 1000 kg/m^3?

Solution:

91.7%

Exercise:

Problem:

Logs sometimes float vertically in a lake because one end has become water-logged and denser than the other. What is the average density of a uniform-diameter log that floats with 20.0% of its length above water?

Exercise:

Problem:

Find the density of a fluid in which a hydrometer having a density of 0.750 g/mL floats with 92.0% of its volume submerged.

Solution:

815 kg/m^3

Exercise:

Problem:

If your body has a density of 995 kg/m^3, what fraction of you will be submerged when floating gently in: (a) freshwater? (b) salt water, which has a density of 1027 kg/m^3?

Exercise:

Problem:

Bird bones have air pockets in them to reduce their weight—this also gives them an average density significantly less than that of the bones of other animals. Suppose an ornithologist weighs a bird bone in air and in water and finds its mass is 45.0 g and its apparent mass when submerged is 3.60 g (the bone is watertight). (a) What mass of water is displaced? (b) What is the volume of the bone? (c) What is its average density?

Solution:

(a) 41.4 g

(b) 41.4 cm^3

(c) 1.09 g/cm^3

Exercise:

Problem:

A rock with a mass of 540 g in air is found to have an apparent mass of 342 g when submerged in water. (a) What mass of water is displaced? (b) What is the volume of the rock? (c) What is its average density? Is this consistent with the value for granite?

Exercise:

Problem:

Archimedes' principle can be used to calculate the density of a fluid as well as that of a solid. Suppose a chunk of iron with a mass of 390.0 g in air is found to have an apparent mass of 350.5 g when completely submerged in an unknown liquid. (a) What mass of fluid does the iron displace? (b) What is the volume of iron, using its density as given in [link] (c) Calculate the fluid's density and identify it.

Solution:

(a) 39.5 g

(b) 50 cm^3

(c) 0.79 g/cm^3

It is ethyl alcohol.

Exercise:

Problem:

In an immersion measurement of a woman's density, she is found to have a mass of 62.0 kg in air and an apparent mass of 0.0850 kg when completely submerged with lungs empty. (a) What mass of water does she displace? (b) What is her volume? (c) Calculate her density. (d) If her lung capacity is 1.75 L, is she able to float without treading water with her lungs filled with air?

Exercise:

Problem:

Some fish have a density slightly less than that of water and must exert a force (swim) to stay submerged. What force must an 85.0-kg grouper exert to stay submerged in salt water if its body density is 1015 kg/m^3 ?

Solution:

8.21 N

Exercise:

Problem:

(a) Calculate the buoyant force on a 2.00-L helium balloon. (b) Given the mass of the rubber in the balloon is 1.50 g, what is the net vertical force on the balloon if it is let go? You can neglect the volume of the rubber.

Exercise:

Problem:

(a) What is the density of a woman who floats in freshwater with 4.00% of her volume above the surface? This could be measured by placing her in a tank with marks on the side to measure how much water she displaces when floating and when held under water (briefly). (b) What percent of her volume is above the surface when she floats in seawater?

Solution:

(a) 960 kg/m^3

(b) 6.34%

She indeed floats more in seawater.

Exercise:

Problem:

A certain man has a mass of 80 kg and a density of $955\ \text{kg/m}^3$ (excluding the air in his lungs). (a) Calculate his volume. (b) Find the buoyant force air exerts on him. (c) What is the ratio of the buoyant force to his weight?

Exercise:

Problem:

A simple compass can be made by placing a small bar magnet on a cork floating in water. (a) What fraction of a plain cork will be submerged when floating in water? (b) If the cork has a mass of 10.0 g and a 20.0-g magnet is placed on it, what fraction of the cork will be submerged? (c) Will the bar magnet and cork float in ethyl alcohol?

Solution:

(a) 0.24

(b) 0.68

(c) Yes, the cork will float because $\rho_{\text{obj}} < \rho_{\text{ethyl alcohol}}(0.678\ \text{g/cm}^3 < 0.79\ \text{g/cm}^3)$

Exercise:

Problem:

What fraction of an iron anchor’s weight will be supported by buoyant force when submerged in saltwater?

Exercise:

Problem:

Scurrilous con artists have been known to represent gold-plated tungsten ingots as pure gold and sell them to the greedy at prices much below gold value but deservedly far above the cost of tungsten. With what accuracy must you be able to measure the mass of such an ingot in and out of water to tell that it is almost pure tungsten rather than pure gold?

Solution:

The difference is 0.006%.

Exercise:

Problem:

A twin-sized air mattress used for camping has dimensions of 100 cm by 200 cm by 15 cm when blown up. The weight of the mattress is 2 kg. How heavy a person could the air mattress hold if it is placed in freshwater?

Exercise:

Problem:

Referring to [link], prove that the buoyant force on the cylinder is equal to the weight of the fluid displaced (Archimedes' principle). You may assume that the buoyant force is $F_2 - F_1$ and that the ends of the cylinder have equal areas A. Note that the volume of the cylinder (and that of the fluid it displaces) equals $(h_2 - h_1)A$.

Solution:

$$F_{\text{net}} = F_2 - F_1 = P_2 A - P_1 A = (P_2 - P_1)A$$

$$= (h_2 \rho_{\text{fl}} g - h_1 \rho_{\text{fl}} g)A$$

$$= (h_2 - h_1)\rho_{\text{fl}}\text{gA}$$

where ρ_{fl} = density of fluid. Therefore,

$$F_{\text{net}} = (h_2 - h_1)A\rho_{\text{fl}}g = V_{\text{fl}}\rho_{\text{fl}}g = m_{\text{fl}}g = w_{\text{fl}}$$

where is w_{fl} the weight of the fluid displaced.

Exercise:

Problem:

(a) A 75.0-kg man floats in freshwater with 3.00% of his volume above water when his lungs are empty, and 5.00% of his volume above water when his lungs are full. Calculate the volume of air he inhales—called his lung capacity—in liters. (b) Does this lung volume seem reasonable?

Glossary

Archimedes’ principle
the buoyant force on an object equals the weight of the fluid it displaces

buoyant force
the net upward force on any object in any fluid

specific gravity
the ratio of the density of an object to a fluid (usually water)

Cohesion and Adhesion in Liquids: Surface Tension and Capillary Action

- Understand cohesive and adhesive forces.
- Define surface tension.
- Understand capillary action.

Cohesion and Adhesion in Liquids

Children blow soap bubbles and play in the spray of a sprinkler on a hot summer day. (See [link].) An underwater spider keeps his air supply in a shiny bubble he carries wrapped around him. A technician draws blood into a small-diameter tube just by touching it to a drop on a pricked finger. A premature infant struggles to inflate her lungs. What is the common thread? All these activities are dominated by the attractive forces between atoms and molecules in liquids—both within a liquid and between the liquid and its surroundings.

Attractive forces between molecules of the same type are called **cohesive forces**. Liquids can, for example, be held in open containers because cohesive forces hold the molecules together. Attractive forces between molecules of different types are called **adhesive forces**. Such forces cause liquid drops to cling to window panes, for example. In this section we examine effects directly attributable to cohesive and adhesive forces in liquids.

Note:
Cohesive Forces
Attractive forces between molecules of the same type are called cohesive forces.

Note:
Adhesive Forces

Attractive forces between molecules of different types are called adhesive forces.

The soap bubbles in this photograph are caused by cohesive forces among molecules in liquids. (credit: Steven Depolo, Flickr)

Surface Tension

Cohesive forces between molecules cause the surface of a liquid to contract to the smallest possible surface area. This general effect is called **surface tension**. Molecules on the surface are pulled inward by cohesive forces, reducing the surface area. Molecules inside the liquid experience zero net force, since they have neighbors on all sides.

Note:
Surface Tension

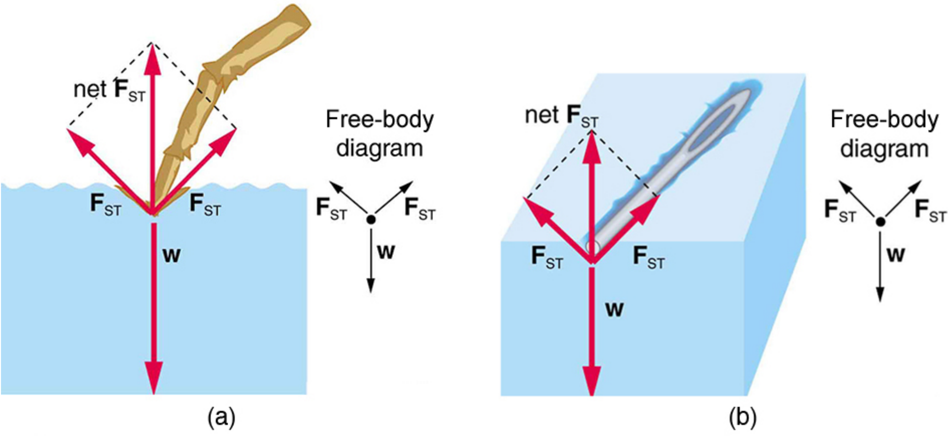
Cohesive forces between molecules cause the surface of a liquid to contract to the smallest possible surface area. This general effect is called surface tension.

Note:

Making Connections: Surface Tension

Forces between atoms and molecules underlie the macroscopic effect called surface tension. These attractive forces pull the molecules closer together and tend to minimize the surface area. This is another example of a submicroscopic explanation for a macroscopic phenomenon.

The model of a liquid surface acting like a stretched elastic sheet can effectively explain surface tension effects. For example, some insects can walk on water (as opposed to floating in it) as we would walk on a trampoline—they dent the surface as shown in [link](a). [link](b) shows another example, where a needle rests on a water surface. The iron needle cannot, and does not, float, because its density is greater than that of water. Rather, its weight is supported by forces in the stretched surface that try to make the surface smaller or flatter. If the needle were placed point down on the surface, its weight acting on a smaller area would break the surface, and it would sink.



Surface tension supporting the weight of an insect and an iron needle, both of which rest on the surface without penetrating it. They are not floating; rather, they are supported by the surface of the liquid. (a) An insect leg dents the water surface. F_{ST} is a restoring force (surface tension) parallel to the surface. (b) An iron needle similarly dents a water surface until the restoring force (surface tension) grows to equal its weight.

Surface tension is proportional to the strength of the cohesive force, which varies with the type of liquid. Surface tension γ is defined to be the force F per unit length L exerted by a stretched liquid membrane:
Equation:

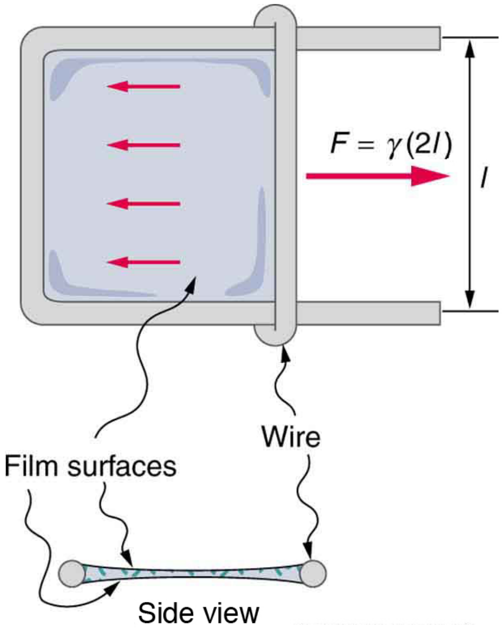
$$\gamma = \frac{F}{L}.$$

[link] lists values of γ for some liquids. For the insect of [link](a), its weight w is supported by the upward components of the surface tension force: $w = \gamma L \sin \theta$, where L is the circumference of the insect’s foot in contact with the water. [link] shows one way to measure surface tension. The liquid film exerts a force on the movable wire in an attempt to reduce its surface area. The magnitude of this force depends on the surface tension of the liquid and can be measured accurately.

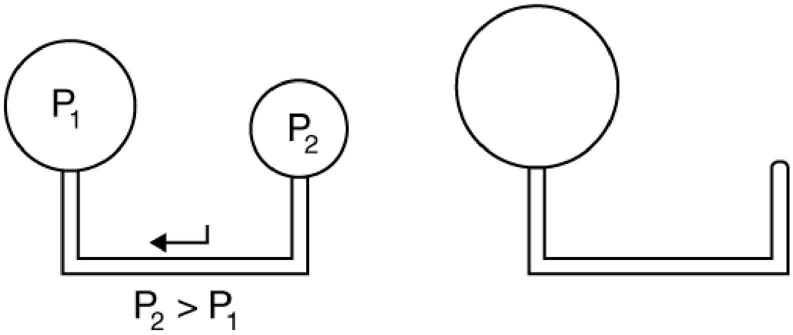
Surface tension is the reason why liquids form bubbles and droplets. The inward surface tension force causes bubbles to be approximately spherical and raises the pressure of the gas trapped inside relative to atmospheric pressure outside. It can be shown that the gauge pressure P inside a spherical bubble is given by
Equation:

$$P = \frac{4\gamma}{r},$$

where r is the radius of the bubble. Thus the pressure inside a bubble is greatest when the bubble is the smallest. Another bit of evidence for this is illustrated in [link]. When air is allowed to flow between two balloons of unequal size, the smaller balloon tends to collapse, filling the larger balloon.



Sliding wire device used for measuring surface tension; the device exerts a force to reduce the film’s surface area. The force needed to hold the wire in place is $F = \gamma L = \gamma(2l)$, since there are *two* liquid surfaces attached to the wire. This force remains nearly constant as the film is stretched, until the film approaches its breaking point.



With the valve closed, two balloons of different sizes are attached to each end of a tube. Upon opening the valve, the smaller balloon decreases in size with the air moving to fill the larger balloon. The pressure in a spherical balloon is inversely proportional to its radius, so that the smaller balloon has a greater internal pressure than the larger balloon, resulting in this flow.

Liquid	**Surface tension γ(N/m)**
Water at 0°C	0.0756

Liquid	**Surface tension γ(N/m)**
Water at 20ºC	0.0728
Water at 100ºC	0.0589
Soapy water (typical)	0.0370
Ethyl alcohol	0.0223
Glycerin	0.0631
Mercury	0.465
Olive oil	0.032
Tissue fluids (typical)	0.050
Blood, whole at 37ºC	0.058
Blood plasma at 37ºC	0.073
Gold at 1070ºC	1.000
Oxygen at −193ºC	0.0157
Helium at −269ºC	0.00012

Surface Tension of Some Liquids[footnote]
At 20ºC unless otherwise stated.

Example:
Surface Tension: Pressure Inside a Bubble
Calculate the gauge pressure inside a soap bubble $2.00 \times 10^{-4}\ \text{m}$ in radius using the surface tension for soapy water in [link]. Convert this pressure to

mm Hg.

Strategy

The radius is given and the surface tension can be found in [link], and so P can be found directly from the equation $P = \frac{4\gamma}{r}$.

Solution

Substituting r and γ into the equation $P = \frac{4\gamma}{r}$, we obtain

Equation:

$$P = \frac{4\gamma}{r} = \frac{4(0.037\text{ N/m})}{2.00 \times 10^{-4}\text{ m}} = 740\text{ N/m}^2 = 740\text{ Pa}.$$

We use a conversion factor to get this into units of mm Hg:

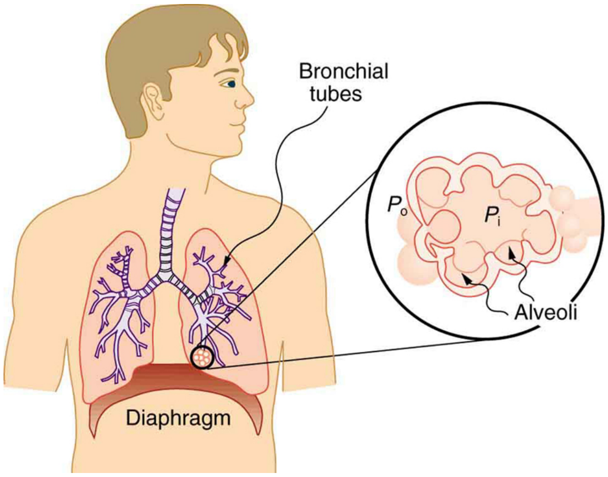
Equation:

$$P = \left(740\text{ N/m}^2\right)\frac{1.00\text{ mm Hg}}{133\text{ N/m}^2} = 5.56\text{ mm Hg}.$$

Discussion

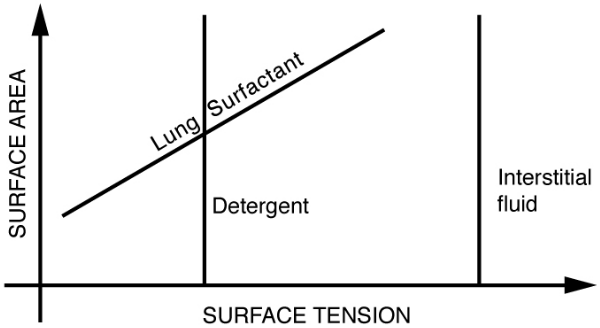
Note that if a hole were to be made in the bubble, the air would be forced out, the bubble would decrease in radius, and the pressure inside would *increase* to atmospheric pressure (760 mm Hg).

Our lungs contain hundreds of millions of mucus-lined sacs called *alveoli*, which are very similar in size, and about 0.1 mm in diameter. (See [link].) You can exhale without muscle action by allowing surface tension to contract these sacs. Medical patients whose breathing is aided by a positive pressure respirator have air blown into the lungs, but are generally allowed to exhale on their own. Even if there is paralysis, surface tension in the alveoli will expel air from the lungs. Since pressure increases as the radii of the alveoli decrease, an occasional deep cleansing breath is needed to fully reinflate the alveoli. Respirators are programmed to do this and we find it natural, as do our companion dogs and cats, to take a cleansing breath before settling into a nap.



Bronchial tubes in the lungs branch into ever-smaller structures, finally ending in alveoli. The alveoli act like tiny bubbles. The surface tension of their mucous lining aids in exhalation and can prevent inhalation if too great.

The tension in the walls of the alveoli results from the membrane tissue and a liquid on the walls of the alveoli containing a long lipoprotein that acts as a surfactant (a surface-tension reducing substance). The need for the surfactant results from the tendency of small alveoli to collapse and the air to fill into the larger alveoli making them even larger (as demonstrated in [link]). During inhalation, the lipoprotein molecules are pulled apart and the wall tension increases as the radius increases (increased surface tension). During exhalation, the molecules slide back together and the surface tension decreases, helping to prevent a collapse of the alveoli. The surfactant therefore serves to change the wall tension so that small alveoli don't collapse and large alveoli are prevented from expanding too much. This tension change is a unique property of these surfactants, and is not shared by detergents (which simply lower surface tension). (See [link].)



Surface tension as a function of surface area. The surface tension for lung surfactant decreases with decreasing area. This ensures that small alveoli don’t collapse and large alveoli are not able to over expand.

If water gets into the lungs, the surface tension is too great and you cannot inhale. This is a severe problem in resuscitating drowning victims. A similar problem occurs in newborn infants who are born without this surfactant—their lungs are very difficult to inflate. This condition is known as *hyaline membrane disease* and is a leading cause of death for infants, particularly in premature births. Some success has been achieved in treating hyaline membrane disease by spraying a surfactant into the infant’s breathing passages. Emphysema produces the opposite problem with alveoli. Alveolar walls of emphysema victims deteriorate, and the sacs combine to form larger sacs. Because pressure produced by surface tension decreases with increasing radius, these larger sacs produce smaller pressure, reducing the ability of emphysema victims to exhale. A common test for emphysema is to measure the pressure and volume of air that can be exhaled.

Note:
Making Connections: Take-Home Investigation

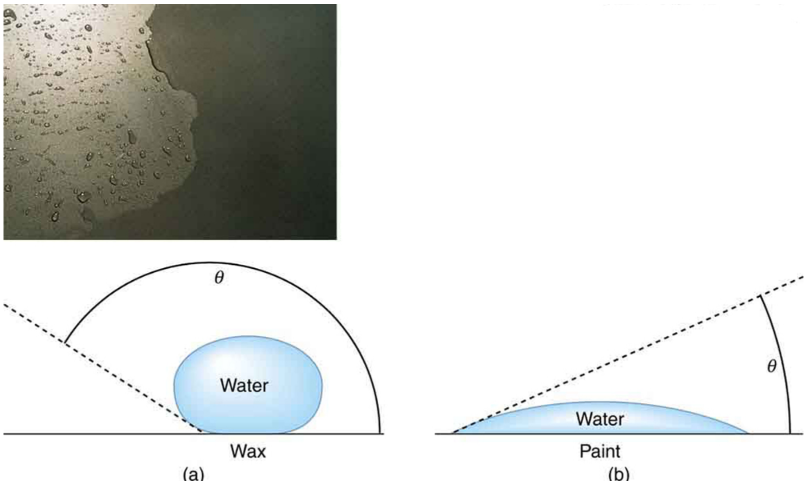
(1) Try floating a sewing needle on water. In order for this activity to work, the needle needs to be very clean as even the oil from your fingers can be sufficient to affect the surface properties of the needle. (2) Place the bristles of a paint brush into water. Pull the brush out and notice that for a short while, the bristles will stick together. The surface tension of the water surrounding the bristles is sufficient to hold the bristles together. As the bristles dry out, the surface tension effect dissipates. (3) Place a loop of thread on the surface of still water in such a way that all of the thread is in contact with the water. Note the shape of the loop. Now place a drop of detergent into the middle of the loop. What happens to the shape of the loop? Why? (4) Sprinkle pepper onto the surface of water. Add a drop of detergent. What happens? Why? (5) Float two matches parallel to each other and add a drop of detergent between them. What happens? Note: For each new experiment, the water needs to be replaced and the bowl washed to free it of any residual detergent.

Adhesion and Capillary Action

Why is it that water beads up on a waxed car but does not on bare paint? The answer is that the adhesive forces between water and wax are much smaller than those between water and paint. Competition between the forces of adhesion and cohesion are important in the macroscopic behavior of liquids. An important factor in studying the roles of these two forces is the angle θ between the tangent to the liquid surface and the surface. (See [link].) The **contact angle** θ is directly related to the relative strength of the cohesive and adhesive forces. The larger the strength of the cohesive force relative to the adhesive force, the larger θ is, and the more the liquid tends to form a droplet. The smaller θ is, the smaller the relative strength, so that the adhesive force is able to flatten the drop. [link] lists contact angles for several combinations of liquids and solids.

Note:
Contact Angle

The angle θ between the tangent to the liquid surface and the surface is called the contact angle.

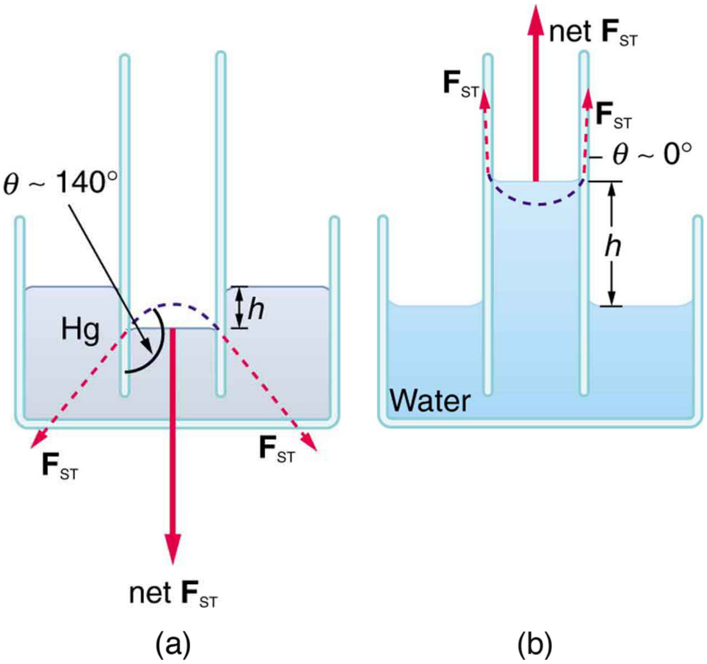


In the photograph, water beads on the waxed car paint and flattens on the unwaxed paint. (a) Water forms beads on the waxed surface because the cohesive forces responsible for surface tension are larger than the adhesive forces, which tend to flatten the drop. (b) Water beads on bare paint are flattened considerably because the adhesive forces between water and paint are strong, overcoming surface tension. The contact angle θ is directly related to the relative strengths of the cohesive and adhesive forces. The larger θ is, the larger the ratio of cohesive to adhesive forces. (credit: P. P. Urone)

One important phenomenon related to the relative strength of cohesive and adhesive forces is **capillary action**—the tendency of a fluid to be raised or suppressed in a narrow tube, or *capillary tube*. This action causes blood to be drawn into a small-diameter tube when the tube touches a drop.

Note:
Capillary Action
The tendency of a fluid to be raised or suppressed in a narrow tube, or capillary tube, is called capillary action.

If a capillary tube is placed vertically into a liquid, as shown in [link], capillary action will raise or suppress the liquid inside the tube depending on the combination of substances. The actual effect depends on the relative strength of the cohesive and adhesive forces and, thus, the contact angle θ given in the table. If θ is less than 90°, then the fluid will be raised; if θ is greater than 90°, it will be suppressed. Mercury, for example, has a very large surface tension and a large contact angle with glass. When placed in a tube, the surface of a column of mercury curves downward, somewhat like a drop. The curved surface of a fluid in a tube is called a **meniscus**. The tendency of surface tension is always to reduce the surface area. Surface tension thus flattens the curved liquid surface in a capillary tube. This results in a downward force in mercury and an upward force in water, as seen in [link].


net $\mathbf{F}_{ST}$
$\mathbf{F}_{ST}$
$\mathbf{F}_{ST}$
$\theta \sim 0°$
h
Water
$\theta \sim 140°$
Hg
h
$\mathbf{F}_{ST}$
$\mathbf{F}_{ST}$
net $\mathbf{F}_{ST}$
(a)
(b)

(a) Mercury is suppressed in a glass tube because its contact angle is greater than 90º. Surface tension exerts a downward force as it flattens the mercury, suppressing it in the tube. The dashed line shows the shape the mercury surface would have without the flattening effect of surface tension. (b) Water is raised in a glass tube because its contact angle is nearly 0º. Surface tension therefore exerts an upward force when it flattens the surface to reduce its area.

Interface	**Contact angle Θ**
Mercury–glass	140º
Water–glass	0º
Water–paraffin	107º
Water–silver	90º
Organic liquids (most)–glass	0º
Ethyl alcohol–glass	0º
Kerosene–glass	26º

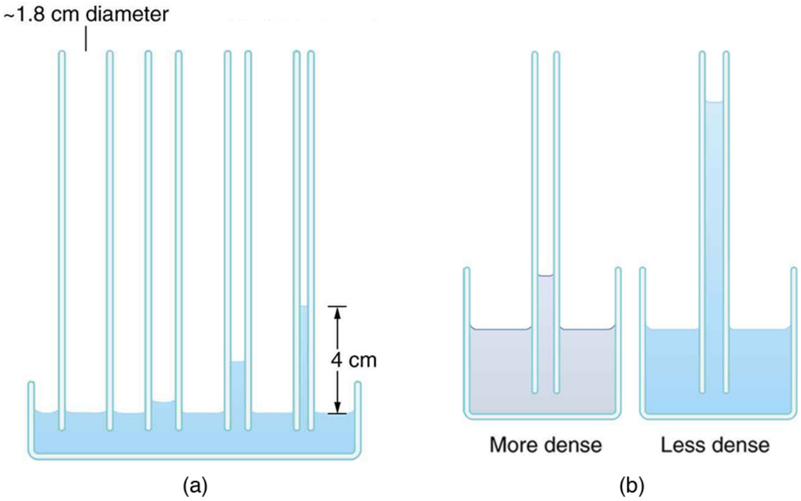
Contact Angles of Some Substances

Capillary action can move liquids horizontally over very large distances, but the height to which it can raise or suppress a liquid in a tube is limited by its weight. It can be shown that this height h is given by

Equation:

$$h = \frac{2\gamma \cos \theta}{\rho \mathrm{gr}}.$$

If we look at the different factors in this expression, we might see how it makes good sense. The height is directly proportional to the surface tension γ, which is its direct cause. Furthermore, the height is inversely proportional to tube radius—the smaller the radius r, the higher the fluid can be raised, since a smaller tube holds less mass. The height is also inversely proportional to fluid density ρ, since a larger density means a greater mass in the same volume. (See [link].)



(a) Capillary action depends on the radius of a tube. The smaller the tube, the greater the height reached. The height is negligible for large-radius tubes. (b) A denser fluid in the same tube rises to a smaller height, all other factors being the same.

Example:
Calculating Radius of a Capillary Tube: Capillary Action: Tree Sap

Can capillary action be solely responsible for sap rising in trees? To answer this question, calculate the radius of a capillary tube that would raise sap 100 m to the top of a giant redwood, assuming that sap's density is $1050\ \text{kg/m}^3$, its contact angle is zero, and its surface tension is the same as that of water at $20.0^{\circ}\ \text{C}$.

Strategy

The height to which a liquid will rise as a result of capillary action is given by $h = \frac{2\gamma\cos\theta}{\rho gr}$, and every quantity is known except for r.

Solution

Solving for r and substituting known values produces

Equation:

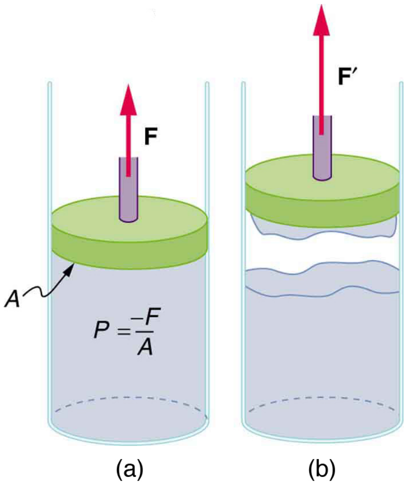
$$\begin{aligned} r &= \frac{2\gamma\cos\theta}{\rho gh} = \frac{2(0.0728\ \text{N/m})\cos(0^{\circ})}{\left(1050\ \text{kg/m}^3\right)\left(9.80\ \text{m/s}^2\right)(100\ \text{m})} \\ &= 1.41 \times 10^{-7}\ \text{m}. \end{aligned}$$

Discussion

This result is unreasonable. Sap in trees moves through the *xylem*, which forms tubes with radii as small as $2.5 \times 10^{-5}\ \text{m}$. This value is about 180 times as large as the radius found necessary here to raise sap $100\ \text{m}$. This means that capillary action alone cannot be solely responsible for sap getting to the tops of trees.

How *does* sap get to the tops of tall trees? (Recall that a column of water can only rise to a height of 10 m when there is a vacuum at the top—see [link].) The question has not been completely resolved, but it appears that it is pulled up like a chain held together by cohesive forces. As each molecule of sap enters a leaf and evaporates (a process called transpiration), the entire chain is pulled up a notch. So a negative pressure created by water evaporation must be present to pull the sap up through the xylem vessels. In most situations, *fluids can push but can exert only negligible pull*, because the cohesive forces seem to be too small to hold the molecules tightly together. But in this case, the cohesive force of water molecules provides a very strong pull. [link] shows one device for studying negative pressure.

Some experiments have demonstrated that negative pressures sufficient to pull sap to the tops of the tallest trees *can* be achieved.



(a) When the piston is raised, it stretches the liquid slightly, putting it under tension and creating a negative absolute pressure $P = -F/A$. (b) The liquid eventually separates, giving an experimental limit to negative pressure in this liquid.

Section Summary

- Attractive forces between molecules of the same type are called cohesive forces.
- Attractive forces between molecules of different types are called adhesive forces.
- Cohesive forces between molecules cause the surface of a liquid to contract to the smallest possible surface area. This general effect is called surface tension.
- Capillary action is the tendency of a fluid to be raised or suppressed in a narrow tube, or capillary tube which is due to the relative strength of cohesive and adhesive forces.

Conceptual Questions

Exercise:

Problem:

The density of oil is less than that of water, yet a loaded oil tanker sits lower in the water than an empty one. Why?

Exercise:

Problem:

Is surface tension due to cohesive or adhesive forces, or both?

Exercise:

Problem:

Is capillary action due to cohesive or adhesive forces, or both?

Exercise:

Problem:

Birds such as ducks, geese, and swans have greater densities than water, yet they are able to sit on its surface. Explain this ability, noting that water does not wet their feathers and that they cannot sit on soapy water.

Exercise:

Problem:

Water beads up on an oily sunbather, but not on her neighbor, whose skin is not oiled. Explain in terms of cohesive and adhesive forces.

Exercise:

Problem:

Could capillary action be used to move fluids in a “weightless” environment, such as in an orbiting space probe?

Exercise:

Problem:

What effect does capillary action have on the reading of a manometer with uniform diameter? Explain your answer.

Exercise:

Problem:

Pressure between the inside chest wall and the outside of the lungs normally remains negative. Explain how pressure inside the lungs can become positive (to cause exhalation) without muscle action.

Problems & Exercises

Exercise:

Problem:

What is the pressure inside an alveolus having a radius of $2.50 \times 10^{-4}\ \text{m}$ if the surface tension of the fluid-lined wall is the same as for soapy water? You may assume the pressure is the same as that created by a spherical bubble.

Solution:

$592\ \text{N/m}^2$

Exercise:

Problem:

(a) The pressure inside an alveolus with a 2.00×10^{-4}-m radius is $1.40 \times 10^3\ \text{Pa}$, due to its fluid-lined walls. Assuming the alveolus acts like a spherical bubble, what is the surface tension of the fluid? (b) Identify the likely fluid. (You may need to extrapolate between values in [link].)

Exercise:

Problem:

What is the gauge pressure in millimeters of mercury inside a soap bubble 0.100 m in diameter?

Solution:

$2.23 \times 10^{-2}\ \text{mm Hg}$

Exercise:

Problem:

Calculate the force on the slide wire in [link] if it is 3.50 cm long and the fluid is ethyl alcohol.

Exercise:

Problem:

[link](a) shows the effect of tube radius on the height to which capillary action can raise a fluid. (a) Calculate the height h for water in a glass tube with a radius of 0.900 cm—a rather large tube like the one on the left. (b) What is the radius of the glass tube on the right if it raises water to 4.00 cm?

Solution:

(a) $1.65 \times 10^{-3}\ \text{m}$

(b) $3.71 \times 10^{-4}\ \text{m}$

Exercise:

Problem:

We stated in [link] that a xylem tube is of radius $2.50 \times 10^{-5}\ \text{m}$. Verify that such a tube raises sap less than a meter by finding h for it, making the same assumptions that sap's density is $1050\ \text{kg/m}^3$, its contact angle is zero, and its surface tension is the same as that of water at $20.0^\text{o}\ \text{C}$.

Exercise:

Problem:

What fluid is in the device shown in [link] if the force is $3.16 \times 10^{-3}\ \text{N}$ and the length of the wire is 2.50 cm? Calculate the surface tension γ and find a likely match from [link].

Solution:

$6.32 \times 10^{-2}\ \text{N/m}$

Based on the values in table, the fluid is probably glycerin.

Exercise:

Problem:

If the gauge pressure inside a rubber balloon with a 10.0-cm radius is 1.50 cm of water, what is the effective surface tension of the balloon?

Exercise:

Problem:

Calculate the gauge pressures inside 2.00-cm-radius bubbles of water, alcohol, and soapy water. Which liquid forms the most stable bubbles, neglecting any effects of evaporation?

Solution:

$$\begin{aligned} P_{\text{w}} &= 14.6\ \text{N/m}^2, \\ P_{\text{a}} &= 4.46\ \text{N/m}^2, \\ P_{\text{sw}} &= 7.40\ \text{N/m}^2. \end{aligned}$$

Alcohol forms the most stable bubble, since the absolute pressure inside is closest to atmospheric pressure.

Exercise:

Problem:

Suppose water is raised by capillary action to a height of 5.00 cm in a glass tube. (a) To what height will it be raised in a paraffin tube of the same radius? (b) In a silver tube of the same radius?

Exercise:

Problem:

Calculate the contact angle θ for olive oil if capillary action raises it to a height of 7.07 cm in a glass tube with a radius of 0.100 mm. Is this value consistent with that for most organic liquids?

Solution:

5.1°

This is near the value of $\theta = 0^{\circ}$ for most organic liquids.

Exercise:

Problem:

When two soap bubbles touch, the larger is inflated by the smaller until they form a single bubble. (a) What is the gauge pressure inside a soap bubble with a 1.50-cm radius? (b) Inside a 4.00-cm-radius soap bubble? (c) Inside the single bubble they form if no air is lost when they touch?

Exercise:

Problem:

Calculate the ratio of the heights to which water and mercury are raised by capillary action in the same glass tube.

Solution:

-2.78

The ratio is negative because water is raised whereas mercury is lowered.

Exercise:

Problem:

What is the ratio of heights to which ethyl alcohol and water are raised by capillary action in the same glass tube?

Glossary

adhesive forces
the attractive forces between molecules of different types

capillary action
the tendency of a fluid to be raised or lowered in a narrow tube

cohesive forces
the attractive forces between molecules of the same type

contact angle
the angle θ between the tangent to the liquid surface and the surface

surface tension
the cohesive forces between molecules which cause the surface of a liquid to contract to the smallest possible surface area

Pressures in the Body

- Explain the concept of pressure the in human body.
- Explain systolic and diastolic blood pressures.
- Describe pressures in the eye, lungs, spinal column, bladder, and skeletal system.

Pressure in the Body

Next to taking a person’s temperature and weight, measuring blood pressure is the most common of all medical examinations. Control of high blood pressure is largely responsible for the significant decreases in heart attack and stroke fatalities achieved in the last three decades. The pressures in various parts of the body can be measured and often provide valuable medical indicators. In this section, we consider a few examples together with some of the physics that accompanies them.

[link] lists some of the measured pressures in mm Hg, the units most commonly quoted.

Body system	**Gauge pressure in mm Hg**
Blood pressures in large arteries (resting)	
Maximum (systolic)	100–140
Minimum (diastolic)	60–90
Blood pressure in large veins	4–15
Eye	12–24
Brain and spinal fluid (lying down)	5–12
Bladder	
While filling	0–25
When full	100–150
Chest cavity between lungs and ribs	−8 to −4
Inside lungs	−2 to +3
Digestive tract	

Body system	**Gauge pressure in mm Hg**
Esophagus	−2
Stomach	0–20
Intestines	10–20
Middle ear	<1

Typical Pressures in Humans

Blood Pressure

Common arterial blood pressure measurements typically produce values of 120 mm Hg and 80 mm Hg, respectively, for systolic and diastolic pressures. Both pressures have health implications. When systolic pressure is chronically high, the risk of stroke and heart attack is increased. If, however, it is too low, fainting is a problem. **Systolic pressure** increases dramatically during exercise to increase blood flow and returns to normal afterward. This change produces no ill effects and, in fact, may be beneficial to the tone of the circulatory system. **Diastolic pressure** can be an indicator of fluid balance. When low, it may indicate that a person is hemorrhaging internally and needs a transfusion. Conversely, high diastolic pressure indicates a ballooning of the blood vessels, which may be due to the transfusion of too much fluid into the circulatory system. High diastolic pressure is also an indication that blood vessels are not dilating properly to pass blood through. This can seriously strain the heart in its attempt to pump blood.

Blood leaves the heart at about 120 mm Hg but its pressure continues to decrease (to almost 0) as it goes from the aorta to smaller arteries to small veins (see [link]). The pressure differences in the circulation system are caused by blood flow through the system as well as the position of the person. For a person standing up, the pressure in the feet will be larger than at the heart due to the weight of the blood $(P = h\rho g)$. If we assume that the distance between the heart and the feet of a person in an upright position is 1.4 m, then the increase in pressure in the feet relative to that in the heart (for a static column of blood) is given by
Equation:

$$\Delta P = \Delta h\rho g = (1.4\ \text{m})\left(1050\ \text{kg/m}^3\right)\left(9.80\ \text{m/s}^2\right) = 1.4 \times 10^4\ \text{Pa} = 108\ \text{mm Hg}.$$

Note:
Increase in Pressure in the Feet of a Person
Equation:

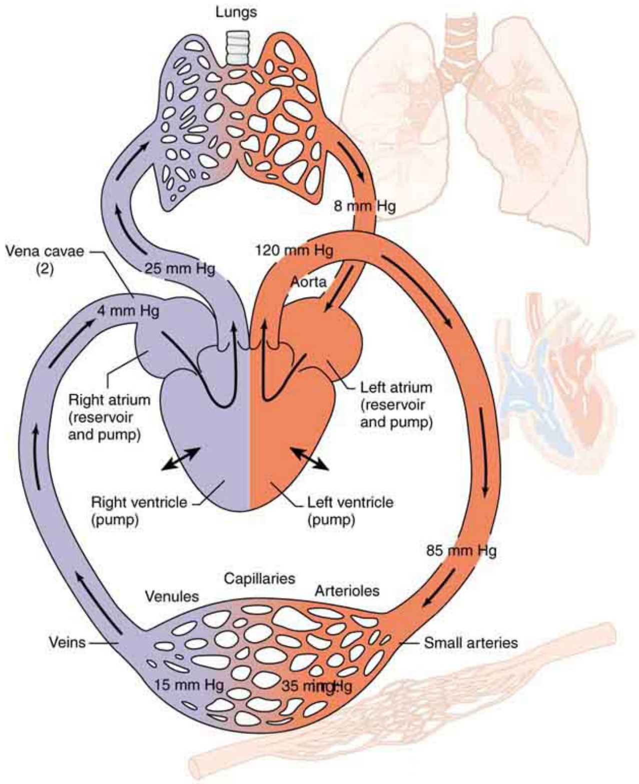
$$\Delta P = \Delta h\rho g = (1.4\ \text{m})\left(1050\ \text{kg/m}^3\right)\left(9.80\ \text{m/s}^2\right) = 1.4 \times 10^4\ \text{Pa} = 108\ \text{mm Hg}.$$

Standing a long time can lead to an accumulation of blood in the legs and swelling. This is the reason why soldiers who are required to stand still for long periods of time have been known to faint. Elastic bandages around the calf can help prevent this accumulation and can also help provide increased pressure to enable the veins to send blood back up to the heart. For similar reasons, doctors recommend tight stockings for long-haul flights.

Blood pressure may also be measured in the major veins, the heart chambers, arteries to the brain, and the lungs. But these pressures are usually only monitored during surgery or for patients in intensive care since the measurements are invasive. To obtain these pressure measurements, qualified health care workers thread thin tubes, called catheters, into appropriate locations to transmit pressures to external measuring devices.

The heart consists of two pumps—the right side forcing blood through the lungs and the left causing blood to flow through the rest of the body ([link]). Right-heart failure, for example, results in a rise in the pressure in the vena cavae and a drop in pressure in the arteries to the lungs. Left-heart failure results in a rise in the pressure entering the left side of the heart and a drop in aortal pressure. Implications of these and other pressures on flow in the circulatory system will be discussed in more detail in Fluid Dynamics and Its Biological and Medical Applications.

Note:
Two Pumps of the Heart
The heart consists of two pumps—the right side forcing blood through the lungs and the left causing blood to flow through the rest of the body.


Lungs
8 mm Hg
Vena cavae (2)
120 mm Hg
25 mm Hg
Aorta
4 mm Hg
Left atrium (reservoir and pump)
Right atrium (reservoir and pump)
Right ventricle (pump)
Left ventricle (pump)
85 mm Hg
Capillaries
Venules
Arterioles
Veins
Small arteries
15 mm Hg
35 mm Hg

Schematic of the circulatory system showing typical pressures. The two pumps in the heart increase pressure and that pressure is reduced as the blood flows through the body. Long-term deviations from these pressures have medical implications discussed in some detail in the Fluid Dynamics and Its Biological and Medical Applications. Only aortal or arterial blood pressure can be measured noninvasively.

Pressure in the Eye

The shape of the eye is maintained by fluid pressure, called **intraocular pressure**, which is normally in the range of 12.0 to 24.0 mm Hg. When the circulation of fluid in the eye is blocked, it can lead to a buildup in pressure, a condition called **glaucoma**. The net pressure can become as great as 85.0 mm Hg, an abnormally large pressure that can permanently damage the optic nerve. To get an idea of the force involved, suppose the back of the eye has an area of 6.0 cm^2, and the net pressure is 85.0 mm Hg. Force is given by $F = \text{PA}$. To get F in newtons, we convert the area to m^2 ($1\text{ m}^2 = 10^4\text{ cm}^2$). Then we calculate as follows:

Equation:

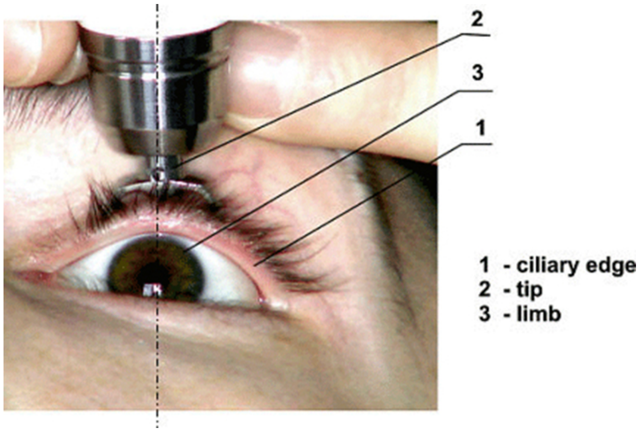
$$F = h\rho g\text{A} = \left(85.0 \times 10^{-3}\ \text{m}\right)\left(13.6 \times 10^{3}\ \text{kg/m}^{3}\right)\left(9.80\ \text{m/s}^{2}\right)\left(6.0 \times 10^{-4}\ \text{m}^{2}\right) = 6.8\ \text{N}.$$

Note:
Eye Pressure
The shape of the eye is maintained by fluid pressure, called intraocular pressure. When the circulation of fluid in the eye is blocked, it can lead to a buildup in pressure, a condition called glaucoma. The force is calculated as
Equation:

$$F = h\rho g\text{A} = \left(85.0 \times 10^{-3}\ \text{m}\right)\left(13.6 \times 10^{3}\ \text{kg/m}^{3}\right)\left(9.80\ \text{m/s}^{2}\right)\left(6.0 \times 10^{-4}\ \text{m}^{2}\right) = 6.8\ \text{N}.$$

This force is the weight of about a 680-g mass. A mass of 680 g resting on the eye (imagine 1.5 lb resting on your eye) would be sufficient to cause it damage. (A normal force here would be the weight of about 120 g, less than one-quarter of our initial value.)

People over 40 years of age are at greatest risk of developing glaucoma and should have their intraocular pressure tested routinely. Most measurements involve exerting a force on the (anesthetized) eye over some area (a pressure) and observing the eye’s response. A noncontact approach uses a puff of air and a measurement is made of the force needed to indent the eye ([link]). If the intraocular pressure is high, the eye will deform less and rebound more vigorously than normal. Excessive intraocular pressures can be detected reliably and sometimes controlled effectively.



The intraocular eye pressure can be read with a tonometer. (credit: DevelopAll at the Wikipedia Project.)

Example:

Calculating Gauge Pressure and Depth: Damage to the Eardrum

Suppose a 3.00-N force can rupture an eardrum. (a) If the eardrum has an area of $1.00\ \text{cm}^2$, calculate the maximum tolerable gauge pressure on the eardrum in newtons per meter squared and convert it to millimeters of mercury. (b) At what depth in freshwater would this person's eardrum rupture, assuming the gauge pressure in the middle ear is zero?

Strategy for (a)

The pressure can be found directly from its definition since we know the force and area. We are looking for the gauge pressure.

Solution for (a)

Equation:

$$P_\text{g} = F/A = 3.00\ \text{N}/(1.00 \times 10^{-4}\ \text{m}^2) = 3.00 \times 10^4\ \text{N/m}^2.$$

We now need to convert this to units of mm Hg:

Equation:

$$P_\text{g} = 3.0 \times 10^4\ \text{N/m}^2 \left(\frac{1.0\ \text{mm Hg}}{133\ \text{N/m}^2}\right) = 226\ \text{mm Hg}.$$

Strategy for (b)

Here we will use the fact that the water pressure varies linearly with depth h below the surface.

Solution for (b)

$P = h\rho g$ and therefore $h = P/\rho g$. Using the value above for P, we have

Equation:

$$h = \frac{3.0 \times 10^4\ \text{N/m}^2}{(1.00 \times 10^3\ \text{kg/m}^3)(9.80\ \text{m/s}^2)} = 3.06\ \text{m}.$$

Discussion

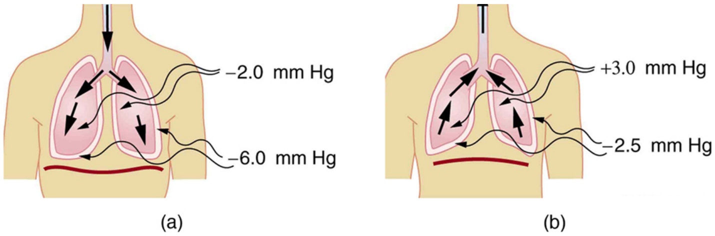
Similarly, increased pressure exerted upon the eardrum from the middle ear can arise when an infection causes a fluid buildup.

Pressure Associated with the Lungs

The pressure inside the lungs increases and decreases with each breath. The pressure drops to below atmospheric pressure (negative gauge pressure) when you inhale, causing air to flow into the lungs. It increases above atmospheric pressure (positive gauge pressure) when you exhale, forcing air out.

Lung pressure is controlled by several mechanisms. Muscle action in the diaphragm and rib cage is necessary for inhalation; this muscle action increases the volume of the lungs thereby reducing the pressure within them [link]. Surface tension in the alveoli creates a positive pressure opposing inhalation. (See Cohesion and Adhesion in Liquids: Surface Tension and Capillary Action.) You can exhale without muscle action by letting surface tension in the alveoli create its own positive pressure. Muscle action can add to this positive pressure to produce forced exhalation, such as when you blow up a balloon, blow out a candle, or cough.

The lungs, in fact, would collapse due to the surface tension in the alveoli, if they were not attached to the inside of the chest wall by liquid adhesion. The gauge pressure in the liquid attaching the lungs to the inside of the chest wall is thus negative, ranging from -4 to -8 mm Hg during exhalation and inhalation, respectively. If air is allowed to enter the chest cavity, it breaks the attachment, and one or both lungs may collapse. Suction is applied to the chest cavity of surgery patients and trauma victims to reestablish negative pressure and inflate the lungs.



(a) During inhalation, muscles expand the chest, and the diaphragm moves downward, reducing pressure inside the lungs to less than atmospheric (negative gauge pressure). Pressure between the lungs and chest wall is even lower to overcome the positive pressure created by surface tension in the lungs. (b) During gentle exhalation, the muscles simply relax and surface tension in the alveoli creates a positive pressure inside the lungs, forcing air out. Pressure between the chest wall and lungs remains negative to keep them attached to the chest wall, but it is less negative than during inhalation.

Other Pressures in the Body

Spinal Column and Skull

Normally, there is a 5- to12-mm Hg pressure in the fluid surrounding the brain and filling the spinal column. This cerebrospinal fluid serves many purposes, one of which is to supply flotation to the brain. The buoyant force supplied by the fluid nearly equals the weight of the brain, since their densities are nearly equal. If there is a loss of fluid, the brain rests on the inside of the skull, causing severe headaches, constricted blood flow, and serious damage. Spinal fluid pressure is measured by means of a needle inserted between vertebrae that transmits the pressure to a suitable measuring device.

Bladder Pressure

This bodily pressure is one of which we are often aware. In fact, there is a relationship between our awareness of this pressure and a subsequent increase in it. Bladder pressure climbs steadily from zero to about 25 mm Hg as the bladder fills to its normal capacity of $500\ \text{cm}^3$. This pressure triggers the **micturition reflex**, which stimulates the feeling of needing to urinate. What is more, it also causes muscles around the bladder to contract, raising the pressure to over 100 mm Hg, accentuating the sensation. Coughing, straining, tensing in cold weather, wearing tight clothes, and experiencing simple nervous tension all can increase bladder pressure and trigger this reflex. So can the weight of a pregnant woman’s fetus, especially if it is kicking vigorously or pushing down with its head! Bladder pressure can be measured by a catheter or by inserting a needle through the bladder wall and transmitting the pressure to an appropriate measuring device. One hazard of high bladder pressure (sometimes created by an obstruction), is that such pressure can force urine back into the kidneys, causing potentially severe damage.

Pressures in the Skeletal System

These pressures are the largest in the body, due both to the high values of initial force, and the small areas to which this force is applied, such as in the joints.. For example, when a person lifts an object improperly, a force of 5000 N may be created between vertebrae in the spine, and this may be applied to an area as small as $10\ \text{cm}^2$. The pressure created is $P = F/A = (5000\ \text{N})/(10^{-3}\ \text{m}^2) = 5.0 \times 10^6\ \text{N/m}^2$ or about 50 atm! This pressure can damage both the spinal discs (the cartilage between vertebrae), as well as the bony vertebrae themselves. Even under normal circumstances, forces between vertebrae in the spine are large enough to create pressures of several atmospheres. Most causes of excessive pressure in the skeletal system can be avoided by lifting properly and avoiding extreme physical activity. (See Forces and Torques in Muscles and Joints.)

There are many other interesting and medically significant pressures in the body. For example, pressure caused by various muscle actions drives food and waste through the digestive system. Stomach pressure behaves much like bladder pressure and is tied to the sensation of hunger. Pressure in the relaxed esophagus is normally negative because pressure in the chest cavity is normally negative. Positive pressure in the stomach may thus force acid into the esophagus, causing “heartburn.” Pressure in the middle ear can result in significant force on the eardrum if it differs greatly from atmospheric pressure, such as while scuba diving. The decrease in external pressure is also noticeable during plane flights (due to a decrease in the weight of air above relative to that at the Earth’s surface). The Eustachian tubes connect the middle ear to the throat and allow us to equalize pressure in the middle ear to avoid an imbalance of force on the eardrum.

Many pressures in the human body are associated with the flow of fluids. Fluid flow will be discussed in detail in the Fluid Dynamics and Its Biological and Medical Applications.

Section Summary

- Measuring blood pressure is among the most common of all medical examinations.
- The pressures in various parts of the body can be measured and often provide valuable medical indicators.
- The shape of the eye is maintained by fluid pressure, called intraocular pressure.
- When the circulation of fluid in the eye is blocked, it can lead to a buildup in pressure, a condition called glaucoma.
- Some of the other pressures in the body are spinal and skull pressures, bladder pressure, pressures in the skeletal system.

Problems & Exercises

Exercise:

Problem:

During forced exhalation, such as when blowing up a balloon, the diaphragm and chest muscles create a pressure of 60.0 mm Hg between the lungs and chest wall. What force in newtons does this pressure create on the $600\ \text{cm}^2$ surface area of the diaphragm?

Solution:

479 N

Exercise:

Problem:

You can chew through very tough objects with your incisors because they exert a large force on the small area of a pointed tooth. What pressure in pascals can you create by exerting a force of $500\ \text{N}$ with your tooth on an area of $1.00\ \text{mm}^2$?

Exercise:

Problem:

One way to force air into an unconscious person’s lungs is to squeeze on a balloon appropriately connected to the subject. What force must you exert on the balloon with your hands to create a gauge pressure of 4.00 cm water, assuming you squeeze on an effective area of $50.0\ \text{cm}^2$?

Solution:

1.96 N

Exercise:

Problem:

Heroes in movies hide beneath water and breathe through a hollow reed (villains never catch on to this trick). In practice, you cannot inhale in this manner if your lungs are more than 60.0 cm below the surface. What is the maximum negative gauge pressure you can create in your lungs on dry land, assuming you can achieve $-3.00\ \text{cm}$ water pressure with your lungs 60.0 cm below the surface?

Solution:

$-63.0\ \text{cm H}_2\text{O}$

Exercise:

Problem:

Gauge pressure in the fluid surrounding an infant's brain may rise as high as 85.0 mm Hg (5 to 12 mm Hg is normal), creating an outward force large enough to make the skull grow abnormally large. (a) Calculate this outward force in newtons on each side of an infant's skull if the effective area of each side is $70.0\ \text{cm}^2$. (b) What is the net force acting on the skull?

Exercise:

Problem:

A full-term fetus typically has a mass of 3.50 kg. (a) What pressure does the weight of such a fetus create if it rests on the mother's bladder, supported on an area of $90.0\ \text{cm}^2$? (b) Convert this pressure to millimeters of mercury and determine if it alone is great enough to trigger the micturition reflex (it will add to any pressure already existing in the bladder).

Solution:

(a) $3.81 \times 10^3\ \text{N/m}^2$

(b) $28.7\ \text{mm Hg}$, which is sufficient to trigger micturition reflex

Exercise:

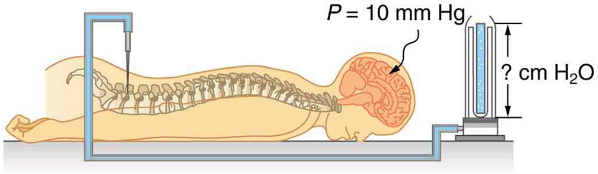
Problem:

If the pressure in the esophagus is $-2.00\ \text{mm Hg}$ while that in the stomach is $+20.0\ \text{mm Hg}$, to what height could stomach fluid rise in the esophagus, assuming a density of 1.10 g/mL? (This movement will not occur if the muscle closing the lower end of the esophagus is working properly.)

Exercise:

Problem:

Pressure in the spinal fluid is measured as shown in [link]. If the pressure in the spinal fluid is 10.0 mm Hg: (a) What is the reading of the water manometer in cm water? (b) What is the reading if the person sits up, placing the top of the fluid 60 cm above the tap? The fluid density is 1.05 g/mL.



A water manometer used to measure pressure in the spinal fluid. The height of the fluid in the manometer is measured relative to the spinal column, and the manometer is open to the atmosphere. The measured pressure will be considerably greater if the person sits up.

Solution:

(a) 13.6 m water

(b) 76.5 cm water

Exercise:

Problem:

Calculate the maximum force in newtons exerted by the blood on an aneurysm, or ballooning, in a major artery, given the maximum blood pressure for this person is 150 mm Hg and the effective area of the aneurysm is $20.0\ \text{cm}^2$. Note that this force is great enough to cause further enlargement and subsequently greater force on the ever-thinner vessel wall.

Exercise:

Problem:

During heavy lifting, a disk between spinal vertebrae is subjected to a 5000-N compressional force. (a) What pressure is created, assuming that the disk has a uniform circular cross section 2.00 cm in radius? (b) What deformation is produced if the disk is 0.800 cm thick and has a Young's modulus of $1.5\times10^{9}\ \text{N/m}^2$?

Solution:

(a) $3.98\times10^{6}\ \text{Pa}$

(b) 2.1×10^{-3} cm

Exercise:

Problem:

When a person sits erect, increasing the vertical position of their brain by 36.0 cm, the heart must continue to pump blood to the brain at the same rate. (a) What is the gain in gravitational potential energy for 100 mL of blood raised 36.0 cm? (b) What is the drop in pressure, neglecting any losses due to friction? (c) Discuss how the gain in gravitational potential energy and the decrease in pressure are related.

Exercise:

Problem:

(a) How high will water rise in a glass capillary tube with a 0.500-mm radius? (b) How much gravitational potential energy does the water gain? (c) Discuss possible sources of this energy.

Solution:

(a) 2.97 cm

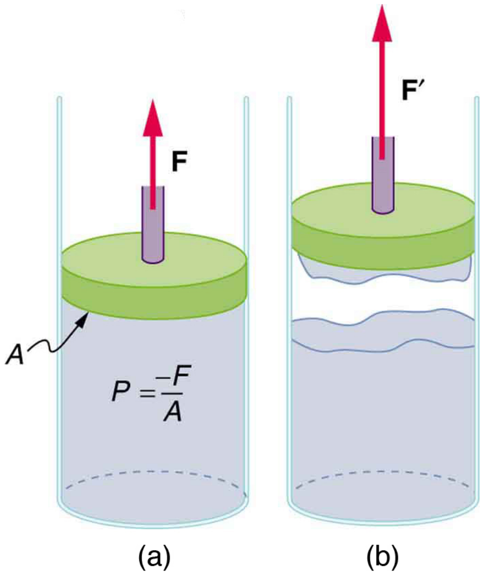
(b) 3.39×10^{-6} J

(c) Work is done by the surface tension force through an effective distance $h/2$ to raise the column of water.

Exercise:

Problem:

A negative pressure of 25.0 atm can sometimes be achieved with the device in [link] before the water separates. (a) To what height could such a negative gauge pressure raise water? (b) How much would a steel wire of the same diameter and length as this capillary stretch if suspended from above?



(a) When the piston is raised, it stretches the liquid slightly, putting it under tension and creating a negative absolute pressure $P = -F/A$ (b) The liquid eventually separates, giving an experimental limit to negative pressure in this liquid.

Exercise:

Problem:

Suppose you hit a steel nail with a 0.500-kg hammer, initially moving at 15.0 m/s and brought to rest in 2.80 mm. (a) What average force is exerted on the nail? (b) How much is the nail compressed if it is 2.50 mm in diameter and 6.00-cm long? (c) What pressure is created on the 1.00-mm-diameter tip of the nail?

Solution:

(a) $2.01 \times 10^{4}\text{ N}$

(b) $1.17 \times 10^{-3}\text{ m}$

(c) $2.56 \times 10^{10}\text{ N/m}^{2}$

Exercise:

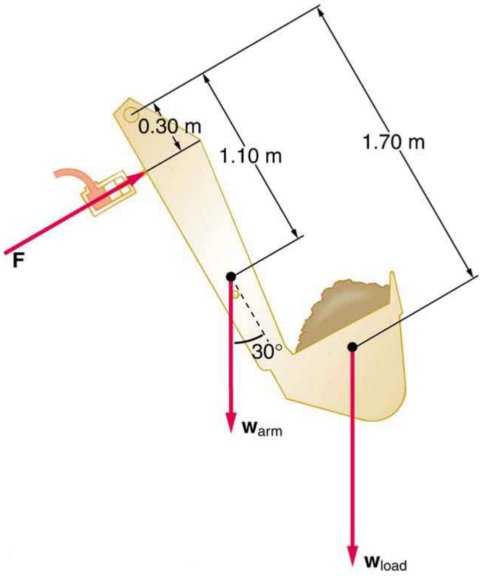
Problem:

Calculate the pressure due to the ocean at the bottom of the Marianas Trench near the Philippines, given its depth is 11.0 km and assuming the density of sea water is constant all the way down. (b) Calculate the percent decrease in volume of sea water due to such a pressure, assuming its bulk modulus is the same as water and is constant. (c) What would be the percent increase in its density? Is the assumption of constant density valid? Will the actual pressure be greater or smaller than that calculated under this assumption?

Exercise:

Problem:

The hydraulic system of a backhoe is used to lift a load as shown in [link]. (a) Calculate the force F the slave cylinder must exert to support the 400-kg load and the 150-kg brace and shovel. (b) What is the pressure in the hydraulic fluid if the slave cylinder is 2.50 cm in diameter? (c) What force would you have to exert on a lever with a mechanical advantage of 5.00 acting on a master cylinder 0.800 cm in diameter to create this pressure?



Hydraulic and mechanical lever systems are used in heavy machinery such as this back hoe.

Solution:

(a) $1.38\times10^{4}\text{ N}$

(b) $2.81\times10^{7}\text{ N/m}^{2}$

(c) 283 N

Exercise:

Problem:

Some miners wish to remove water from a mine shaft. A pipe is lowered to the water 90 m below, and a negative pressure is applied to raise the water. (a) Calculate the pressure needed to raise the water. (b) What is unreasonable about this pressure? (c) What is unreasonable about the premise?

Exercise:

Problem:

You are pumping up a bicycle tire with a hand pump, the piston of which has a 2.00-cm radius.

(a) What force in newtons must you exert to create a pressure of $6.90 \times 10^5 \text{ Pa}$ (b) What is unreasonable about this (a) result? (c) Which premises are unreasonable or inconsistent?

Solution:

(a) 867 N

(b) This is too much force to exert with a hand pump.

(c) The assumed radius of the pump is too large; it would be nearly two inches in diameter—too large for a pump or even a master cylinder. The pressure is reasonable for bicycle tires.

Exercise:

Problem:

Consider a group of people trying to stay afloat after their boat strikes a log in a lake. Construct a problem in which you calculate the number of people that can cling to the log and keep their heads out of the water. Among the variables to be considered are the size and density of the log, and what is needed to keep a person’s head and arms above water without swimming or treading water.

Exercise:

Problem:

The alveoli in emphysema victims are damaged and effectively form larger sacs. Construct a problem in which you calculate the loss of pressure due to surface tension in the alveoli because of their larger average diameters. (Part of the lung’s ability to expel air results from pressure created by surface tension in the alveoli.) Among the things to consider are the normal surface tension of the fluid lining the alveoli, the average alveolar radius in normal individuals and its average in emphysema sufferers.

Glossary

diastolic pressure
minimum arterial blood pressure; indicator for the fluid balance

glaucoma
condition caused by the buildup of fluid pressure in the eye

intraocular pressure
fluid pressure in the eye

micturition reflex
stimulates the feeling of needing to urinate, triggered by bladder pressure

systolic pressure
maximum arterial blood pressure; indicator for the blood flow

Introduction to Fluid Dynamics and Its Biological and Medical Applications

class="introduction"

Many fluids are flowing in this scene. Water from the hose and smoke from the fire are visible flows. Less visible are the flow of air and the flow of fluids on the ground and within the people fighting the fire. Explore all types of flow, such as visible, implied, turbulent, laminar, and so on, present in this scene. Make a list and discuss the relative energies involved in the various flows, including the level of confidence in your estimates. (credit: Andrew Magill, Flickr)

We have dealt with many situations in which fluids are static. But by their very definition, fluids flow. Examples come easily—a column of smoke rises from a camp fire, water streams from a fire hose, blood courses through your veins. Why does rising smoke curl and twist? How does a nozzle increase the speed of water emerging from a hose? How does the body regulate blood flow? The physics of fluids in motion—**fluid dynamics**—allows us to answer these and many other questions.

Glossary

fluid dynamics
: the physics of fluids in motion

Flow Rate and Its Relation to Velocity

- Calculate flow rate.
- Define units of volume.
- Describe incompressible fluids.
- Explain the consequences of the equation of continuity.

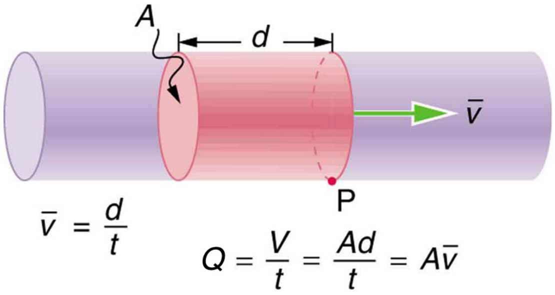
Flow rate Q is defined to be the volume of fluid passing by some location through an area during a period of time, as seen in [link]. In symbols, this can be written as

Equation:

$$Q=\frac{V}{t},$$

where V is the volume and t is the elapsed time.

The SI unit for flow rate is $\text{m}^3\text{/s}$, but a number of other units for Q are in common use. For example, the heart of a resting adult pumps blood at a rate of 5.00 liters per minute (L/min). Note that a **liter** (L) is 1/1000 of a cubic meter or 1000 cubic centimeters ($10^{-3}\ \text{m}^3$ or $10^3\ \text{cm}^3$). In this text we shall use whatever metric units are most convenient for a given situation.



Flow rate is the volume of fluid per unit time flowing past a point through the area A. Here the shaded cylinder of fluid flows past point P in a uniform pipe in time t. The volume of the cylinder is Ad

and the average velocity is
$v = d/t$ so that the flow rate
is $Q = \text{Ad}/t = Av$.

Example:
Calculating Volume from Flow Rate: The Heart Pumps a Lot of Blood in a Lifetime

How many cubic meters of blood does the heart pump in a 75-year lifetime, assuming the average flow rate is 5.00 L/min?

Strategy

Time and flow rate Q are given, and so the volume V can be calculated from the definition of flow rate.

Solution

Solving $Q = V/t$ for volume gives

Equation:

$$V = \text{Qt}.$$

Substituting known values yields

Equation:

$$\begin{aligned} V &= \left(\frac{5.00\ \text{L}}{1\ \text{min}}\right)(75\ \text{y})\left(\frac{1\ \text{m}^3}{10^3\ \text{L}}\right)\left(5.26 \times 10^5\ \frac{\text{min}}{\text{y}}\right) \\ &= 2.0 \times 10^5\ \text{m}^3. \end{aligned}$$

Discussion

This amount is about 200,000 tons of blood. For comparison, this value is equivalent to about 200 times the volume of water contained in a 6-lane 50-m lap pool.

Flow rate and velocity are related, but quite different, physical quantities. To make the distinction clear, think about the flow rate of a river. The greater the velocity of the water, the greater the flow rate of the river. But flow rate also depends on the size of the river. A rapid mountain stream carries far less water than the Amazon River in Brazil, for example. The precise relationship between flow rate Q and velocity v is

Equation:

$$Q = Av,$$

where A is the cross-sectional area and v is the average velocity. This equation seems logical enough. The relationship tells us that flow rate is directly proportional to both the magnitude of the average velocity (hereafter referred to as the speed) and the size of a river, pipe, or other conduit. The larger the conduit, the greater its cross-sectional area. [link] illustrates how this relationship is obtained. The shaded cylinder has a volume

Equation:

$$V = \text{Ad},$$

which flows past the point P in a time t. Dividing both sides of this relationship by t gives

Equation:

$$\frac{V}{t} = \frac{\text{Ad}}{t}.$$

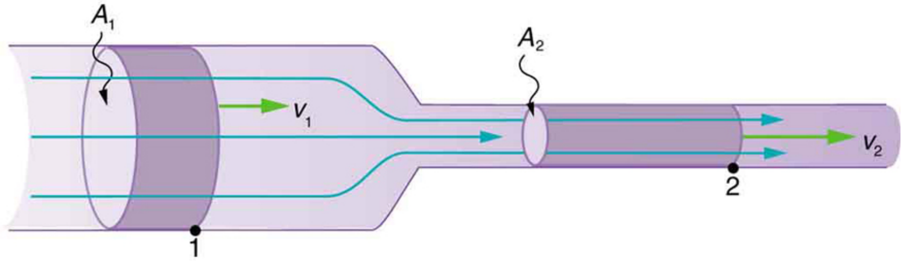
We note that $Q = V/t$ and the average speed is $v = d/t$. Thus the equation becomes $Q = Av$.

[link] shows an incompressible fluid flowing along a pipe of decreasing radius. Because the fluid is incompressible, the same amount of fluid must flow past any point in the tube in a given time to ensure continuity of flow. In this case, because the cross-sectional area of the pipe decreases, the velocity must necessarily increase. This logic can be extended to say that the flow rate must be the same at all points along the pipe. In particular, for points 1 and 2,

Equation:

$$\left.\begin{array}{r}Q_1 = Q_2\\ A_1 v_1 = A_2 v_2\end{array}\right\}.$$

This is called the equation of continuity and is valid for any incompressible fluid. The consequences of the equation of continuity can be observed when water flows from a hose into a narrow spray nozzle: it emerges with a large speed—that is the purpose of the nozzle. Conversely, when a river empties into one end of a reservoir, the water slows considerably, perhaps picking up speed again when it leaves the other end of the reservoir. In other words, speed increases when cross-sectional area decreases, and speed decreases when cross-sectional area increases.



When a tube narrows, the same volume occupies a greater length. For the same volume to pass points 1 and 2 in a given time, the speed must be greater at point 2. The process is exactly reversible. If the fluid flows in the opposite direction, its speed will decrease when the tube widens. (Note that the relative volumes of the two cylinders and the corresponding velocity vector arrows are not drawn to scale.)

Since liquids are essentially incompressible, the equation of continuity is valid for all liquids. However, gases are compressible, and so the equation must be applied with caution to gases if they are subjected to compression or expansion.

Example:

Calculating Fluid Speed: Speed Increases When a Tube Narrows

A nozzle with a radius of 0.250 cm is attached to a garden hose with a radius of 0.900 cm. The flow rate through hose and nozzle is 0.500 L/s. Calculate the speed of the water (a) in the hose and (b) in the nozzle.

Strategy

We can use the relationship between flow rate and speed to find both velocities. We will use the subscript 1 for the hose and 2 for the nozzle.

Solution for (a)

First, we solve $Q = Av$ for v_1 and note that the cross-sectional area is $A = \pi r^2$, yielding

Equation:

$$v_1 = \frac{Q}{A_1} = \frac{Q}{\pi r_1^2}.$$

Substituting known values and making appropriate unit conversions yields

Equation:

$$v_1 = \frac{(0.500\ \text{L/s})(10^{-3}\ \text{m}^3/\text{L})}{\pi(9.00 \times 10^{-3}\ \text{m})^2} = 1.96\ \text{m/s}.$$

Solution for (b)

We could repeat this calculation to find the speed in the nozzle v_2, but we will use the equation of continuity to give a somewhat different insight. Using the equation which states

Equation:

$$A_1 v_1 = A_2 v_2,$$

solving for v_2 and substituting πr^2 for the cross-sectional area yields

Equation:

$$v_2 = \frac{A_1}{A_2} v_1 = \frac{\pi r_1^2}{\pi r_2^2} v_1 = \frac{r_{1^2}}{r_{2^2}} v_1.$$

Substituting known values,

Equation:

$$v_2 = \frac{(0.900\text{ cm})^2}{(0.250\text{ cm})^2} 1.96\text{ m/s} = 25.5\text{ m/s}.$$

Discussion

A speed of 1.96 m/s is about right for water emerging from a nozzleless hose. The nozzle produces a considerably faster stream merely by constricting the flow to a narrower tube.

The solution to the last part of the example shows that speed is inversely proportional to the *square* of the radius of the tube, making for large effects when radius varies. We can blow out a candle at quite a distance, for example, by pursing our lips, whereas blowing on a candle with our mouth wide open is quite ineffective.

In many situations, including in the cardiovascular system, branching of the flow occurs. The blood is pumped from the heart into arteries that subdivide into smaller arteries (arterioles) which branch into very fine vessels called capillaries. In this situation, continuity of flow is maintained but it is the *sum* of the flow rates in each of the branches in any portion along the tube that is maintained. The equation of continuity in a more general form becomes

Equation:

$$n_1 A_1 v_1 = n_2 A_2 v_2,$$

where n_1 and n_2 are the number of branches in each of the sections along the tube.

Example:

Calculating Flow Speed and Vessel Diameter: Branching in the Cardiovascular System

The aorta is the principal blood vessel through which blood leaves the heart in order to circulate around the body. (a) Calculate the average speed of the blood in the aorta if the flow rate is 5.0 L/min. The aorta has a radius of 10 mm. (b) Blood also flows through smaller blood vessels known as capillaries. When the rate of blood flow in the aorta is 5.0 L/min, the speed of blood in the capillaries is about 0.33 mm/s. Given that the average diameter of a capillary is 8.0 μm, calculate the number of capillaries in the blood circulatory system.

Strategy

We can use $Q = Av$ to calculate the speed of flow in the aorta and then use the general form of the equation of continuity to calculate the number of capillaries as all of the other variables are known.

Solution for (a)

The flow rate is given by $Q = Av$ or $v = \frac{Q}{\pi r^2}$ for a cylindrical vessel. Substituting the known values (converted to units of meters and seconds) gives

Equation:

$$v = \frac{(5.0\ \text{L/min})(10^{-3}\ \text{m}^3/\text{L})(1\ \text{min}/60\ \text{s})}{\pi(0.010\ \text{m})^2} = 0.27\ \text{m/s}.$$

Solution for (b)

Using $n_1A_1v_1 = n_2A_2v_1$, assigning the subscript 1 to the aorta and 2 to the capillaries, and solving for n_2 (the number of capillaries) gives $n_2 = \frac{n_1A_1v_1}{A_2v_2}$. Converting all quantities to units of meters and seconds and substituting into the equation above gives

Equation:

$$n_2 = \frac{(1)(\pi)\left(10 \times 10^{-3}\ \text{m}\right)^2(0.27\ \text{m/s})}{(\pi)\left(4.0 \times 10^{-6}\ \text{m}\right)^2\left(0.33 \times 10^{-3}\ \text{m/s}\right)} = 5.0 \times 10^9\ \text{capillaries}.$$

Discussion

Note that the speed of flow in the capillaries is considerably reduced relative to the speed in the aorta due to the significant increase in the total cross-sectional area at the capillaries. This low speed is to allow sufficient time for effective exchange to occur although it is equally important for the flow not to become stationary in order to avoid the possibility of clotting. Does this large number of capillaries in the body seem reasonable? In active muscle, one finds about 200 capillaries per mm^3, or about 200×10^6 per 1 kg of muscle. For 20 kg of muscle, this amounts to about 4×10^9 capillaries.

Section Summary

- Flow rate Q is defined to be the volume V flowing past a point in time t, or $Q=\frac{V}{t}$ where V is volume and t is time.
- The SI unit of volume is m^3.
- Another common unit is the liter (L), which is $10^{-3}\ \text{m}^3$.
- Flow rate and velocity are related by $Q=Av$ where A is the cross-sectional area of the flow and v is its average velocity.
- For incompressible fluids, flow rate at various points is constant. That is,

  Equation:

$$\begin{array}{rcl} Q_1 & = & Q_2 \\ A_1v_1 & = & A_2v_2 \\ n_1A_1v_1 & = & n_2A_2v_2 \end{array}\qquad .$$

Conceptual Questions

Exercise:

Problem:

What is the difference between flow rate and fluid velocity? How are they related?

Exercise:

Problem:

Many figures in the text show streamlines. Explain why fluid velocity is greatest where streamlines are closest together. (Hint: Consider the relationship between fluid velocity and the cross-sectional area through which it flows.)

Exercise:

Problem:

Identify some substances that are incompressible and some that are not.

Problems & Exercises

Exercise:

Problem:

What is the average flow rate in cm^3/s of gasoline to the engine of a car traveling at 100 km/h if it averages 10.0 km/L?

Solution:

$2.78\ \text{cm}^3/\text{s}$

Exercise:

Problem:

The heart of a resting adult pumps blood at a rate of 5.00 L/min. (a) Convert this to cm^3/s. (b) What is this rate in m^3/s?

Exercise:

Problem:

Blood is pumped from the heart at a rate of 5.0 L/min into the aorta (of radius 1.0 cm). Determine the speed of blood through the aorta.

Solution:

27 cm/s

Exercise:

Problem:

Blood is flowing through an artery of radius 2 mm at a rate of 40 cm/s. Determine the flow rate and the volume that passes through the artery in a period of 30 s.

Exercise:

Problem:

The Huka Falls on the Waikato River is one of New Zealand’s most visited natural tourist attractions (see [link]). On average the river has a flow rate of about 300,000 L/s. At the gorge, the river narrows to 20 m wide and averages 20 m deep. (a) What is the average speed of the river in the gorge? (b) What is the average speed of the water in the river downstream of the falls when it widens to 60 m and its depth increases to an average of 40 m?

The Huka Falls in Taupo, New Zealand, demonstrate flow rate. (credit: RaviGogna, Flickr)

Solution:

(a) 0.75 m/s

(b) 0.13 m/s

Exercise:

Problem:

A major artery with a cross-sectional area of 1.00 cm^2 branches into 18 smaller arteries, each with an average cross-sectional area of 0.400 cm^2. By what factor is the average velocity of the blood reduced when it passes into these branches?

Exercise:

Problem:

(a) As blood passes through the capillary bed in an organ, the capillaries join to form venules (small veins). If the blood speed increases by a factor of 4.00 and the total cross-sectional area of the venules is 10.0 cm^2, what is the total cross-sectional area of the capillaries feeding these venules? (b) How many capillaries are involved if their average diameter is $10.0 \text{ }\mu\text{m}$?

Solution:

(a) 40.0 cm^2

(b) 5.09×10^7

Exercise:

Problem:

The human circulation system has approximately 1×10^9 capillary vessels. Each vessel has a diameter of about $8 \text{ }\mu\text{m}$. Assuming cardiac output is 5 L/min, determine the average velocity of blood flow through each capillary vessel.

Exercise:

Problem:

(a) Estimate the time it would take to fill a private swimming pool with a capacity of 80,000 L using a garden hose delivering 60 L/min. (b) How long would it take to fill if you could divert a moderate size river, flowing at $5000\ \text{m}^3/\text{s}$, into it?

Solution:

(a) 22 h

(b) 0.016 s

Exercise:

Problem:

The flow rate of blood through a $2.00 \times 10^{-6}\text{-m}$ -radius capillary is $3.80 \times 10^{-9}\ \text{cm}^3/\text{s}$. (a) What is the speed of the blood flow? (This small speed allows time for diffusion of materials to and from the blood.) (b) Assuming all the blood in the body passes through capillaries, how many of them must there be to carry a total flow of $90.0\ \text{cm}^3/\text{s}$? (The large number obtained is an overestimate, but it is still reasonable.)

Exercise:

Problem:

(a) What is the fluid speed in a fire hose with a 9.00-cm diameter carrying 80.0 L of water per second? (b) What is the flow rate in cubic meters per second? (c) Would your answers be different if salt water replaced the fresh water in the fire hose?

Solution:

(a) 12.6 m/s

(b) $0.0800 \text{ m}^3/\text{s}$

(c) No, independent of density.

Exercise:

Problem:

The main uptake air duct of a forced air gas heater is 0.300 m in diameter. What is the average speed of air in the duct if it carries a volume equal to that of the house's interior every 15 min? The inside volume of the house is equivalent to a rectangular solid 13.0 m wide by 20.0 m long by 2.75 m high.

Exercise:

Problem:

Water is moving at a velocity of 2.00 m/s through a hose with an internal diameter of 1.60 cm. (a) What is the flow rate in liters per second? (b) The fluid velocity in this hose's nozzle is 15.0 m/s. What is the nozzle's inside diameter?

Solution:

(a) 0.402 L/s

(b) 0.584 cm

Exercise:

Problem:

Prove that the speed of an incompressible fluid through a constriction, such as in a Venturi tube, increases by a factor equal to the square of the factor by which the diameter decreases. (The converse applies for flow out of a constriction into a larger-diameter region.)

Exercise:

Problem:

Water emerges straight down from a faucet with a 1.80-cm diameter at a speed of 0.500 m/s. (Because of the construction of the faucet, there is no variation in speed across the stream.) (a) What is the flow rate in cm^3/s? (b) What is the diameter of the stream 0.200 m below the faucet? Neglect any effects due to surface tension.

Solution:

(a) $127\ \text{cm}^3/\text{s}$

(b) 0.890 cm

Exercise:

Problem: **Unreasonable Results**

A mountain stream is 10.0 m wide and averages 2.00 m in depth. During the spring runoff, the flow in the stream reaches $100{,}000\ \text{m}^3/\text{s}$. (a) What is the average velocity of the stream under these conditions? (b) What is unreasonable about this velocity? (c) What is unreasonable or inconsistent about the premises?

Glossary

flow rate	abbreviated Q, it is the volume V that flows past a particular point during a time t, or $Q = V/t$
liter	a unit of volume, equal to $10^{-3}\ \text{m}^3$

Bernoulli’s Equation

- Explain the terms in Bernoulli’s equation.
- Explain how Bernoulli’s equation is related to conservation of energy.
- Explain how to derive Bernoulli’s principle from Bernoulli’s equation.
- Calculate with Bernoulli’s principle.
- List some applications of Bernoulli’s principle.

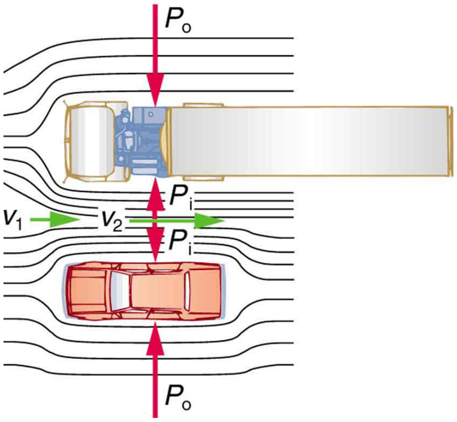
When a fluid flows into a narrower channel, its speed increases. That means its kinetic energy also increases. Where does that change in kinetic energy come from? The increased kinetic energy comes from the net work done on the fluid to push it into the channel and the work done on the fluid by the gravitational force, if the fluid changes vertical position. Recall the work-energy theorem,

Equation:

$$W_{\text{net}} = \frac{1}{2}\text{mv}^2 - \frac{1}{2}\text{mv}_0^2.$$

There is a pressure difference when the channel narrows. This pressure difference results in a net force on the fluid: recall that pressure times area equals force. The net work done increases the fluid’s kinetic energy. As a result, the *pressure will drop in a rapidly-moving fluid*, whether or not the fluid is confined to a tube.

There are a number of common examples of pressure dropping in rapidly-moving fluids. Shower curtains have a disagreeable habit of bulging into the shower stall when the shower is on. The high-velocity stream of water and air creates a region of lower pressure inside the shower, and standard atmospheric pressure on the other side. The pressure difference results in a net force inward pushing the curtain in. You may also have noticed that when passing a truck on the highway, your car tends to veer toward it. The reason is the same—the high velocity of the air between the car and the truck creates a region of lower pressure, and the vehicles are pushed together by greater pressure on the outside. (See [link].) This effect was observed as far back as the mid-1800s, when it was found that trains passing in opposite directions tipped precariously toward one another.



An overhead view of a car passing a truck on a highway. Air passing between the vehicles flows in a narrower channel and must increase its speed (v_2 is greater than v_1), causing the pressure between them to drop (P_i is less than P_o). Greater pressure on the outside pushes the car and truck together.

Note:

Making Connections: Take-Home Investigation with a Sheet of Paper

Hold the short edge of a sheet of paper parallel to your mouth with one hand on each side of your mouth. The page should slant downward over your hands. Blow over the top of the page. Describe what happens and explain the reason for this behavior.

Bernoulli’s Equation

The relationship between pressure and velocity in fluids is described quantitatively by **Bernoulli’s equation**, named after its discoverer, the Swiss scientist Daniel Bernoulli (1700–1782). Bernoulli’s equation states that for an incompressible, frictionless fluid, the following sum is constant:

Equation:

$$P+\frac{1}{2}\rho v^2+\rho \text{gh}=\text{constant},$$

where P is the absolute pressure, ρ is the fluid density, v is the velocity of the fluid, h is the height above some reference point, and g is the acceleration due to gravity. If we follow a small volume of fluid along its path, various quantities in the sum may change, but the total remains constant. Let the subscripts 1 and 2 refer to any two points along the path that the bit of fluid follows; Bernoulli’s equation becomes

Equation:

$$P_1+\frac{1}{2}\rho v_1^2+\rho\, gh_1=P_2+\frac{1}{2}\rho v_2^2+\rho\, gh_2\,.$$

Bernoulli’s equation is a form of the conservation of energy principle. Note that the second and third terms are the kinetic and potential energy with m replaced by ρ. In fact, each term in the equation has units of energy per unit volume. We can prove this for the second term by substituting $\rho=m/V$ into it and gathering terms:

Equation:

$$\frac{1}{2}\rho v^2=\frac{\frac{1}{2}\text{mv}^2}{V}=\frac{\text{KE}}{V}.$$

So $\frac{1}{2}\rho v^2$ is the kinetic energy per unit volume. Making the same substitution into the third term in the equation, we find

Equation:

$$\rho\, gh = \frac{mgh}{V} = \frac{\mathrm{PE_g}}{V}\,,$$

so $\rho\mathrm{gh}$ is the gravitational potential energy per unit volume. Note that pressure P has units of energy per unit volume, too. Since $P = F/A$, its units are $\mathrm{N/m^2}$. If we multiply these by m/m, we obtain $\mathrm{N \cdot m/m^3 = J/m^3}$, or energy per unit volume. Bernoulli's equation is, in fact, just a convenient statement of conservation of energy for an incompressible fluid in the absence of friction.

Note:
Making Connections: Conservation of Energy
Conservation of energy applied to fluid flow produces Bernoulli's equation. The net work done by the fluid's pressure results in changes in the fluid's KE and $\mathrm{PE_g}$ per unit volume. If other forms of energy are involved in fluid flow, Bernoulli's equation can be modified to take these forms into account. Such forms of energy include thermal energy dissipated because of fluid viscosity.

The general form of Bernoulli's equation has three terms in it, and it is broadly applicable. To understand it better, we will look at a number of specific situations that simplify and illustrate its use and meaning.

Bernoulli's Equation for Static Fluids

Let us first consider the very simple situation where the fluid is static—that is, $v_1 = v_2 = 0$. Bernoulli's equation in that case is
Equation:

$$P_1 + \rho\, gh_1 = P_2 + \rho\, gh_2\,.$$

We can further simplify the equation by taking $h_2 = 0$ (we can always choose some height to be zero, just as we often have done for other situations involving the gravitational force, and take all other heights to be relative to this). In that case, we get

Equation:

$$P_2 = P_1 + \rho\, g h_1 \,.$$

This equation tells us that, in static fluids, pressure increases with depth. As we go from point 1 to point 2 in the fluid, the depth increases by h_1, and consequently, P_2 is greater than P_1 by an amount $\rho\, g h_1$. In the very simplest case, P_1 is zero at the top of the fluid, and we get the familiar relationship $P = \rho\, g h$. (Recall that $P = \rho g h$ and $\Delta \text{PE}_{\text{g}} = \text{mgh}$.) Bernoulli's equation includes the fact that the pressure due to the weight of a fluid is ρgh. Although we introduce Bernoulli's equation for fluid flow, it includes much of what we studied for static fluids in the preceding chapter.

Bernoulli's Principle—Bernoulli's Equation at Constant Depth

Another important situation is one in which the fluid moves but its depth is constant—that is, $h_1 = h_2$. Under that condition, Bernoulli's equation becomes

Equation:

$$P_1 + \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho v_2^2.$$

Situations in which fluid flows at a constant depth are so important that this equation is often called **Bernoulli's principle**. It is Bernoulli's equation for fluids at constant depth. (Note again that this applies to a small volume of fluid as we follow it along its path.) As we have just discussed, pressure drops as speed increases in a moving fluid. We can see this from Bernoulli's principle. For example, if v_2 is greater than v_1 in the equation, then P_2 must be less than P_1 for the equality to hold.

Example:

Calculating Pressure: Pressure Drops as a Fluid Speeds Up

In [link], we found that the speed of water in a hose increased from 1.96 m/s to 25.5 m/s going from the hose to the nozzle. Calculate the pressure in the hose, given that the absolute pressure in the nozzle is $1.01 \times 10^5 \text{ N/m}^2$ (atmospheric, as it must be) and assuming level, frictionless flow.

Strategy

Level flow means constant depth, so Bernoulli's principle applies. We use the subscript 1 for values in the hose and 2 for those in the nozzle. We are thus asked to find P_1.

Solution

Solving Bernoulli's principle for P_1 yields

Equation:

$$P_1 = P_2 + \frac{1}{2}\rho v_2^2 - \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho(v_2^2 - v_1^2).$$

Substituting known values,

Equation:

$$\begin{aligned} P_1 &= 1.01 \times 10^5 \text{ N/m}^2 \\ &\quad + \tfrac{1}{2}(10^3 \text{ kg/m}^3)\left[(25.5 \text{ m/s})^2 - (1.96 \text{ m/s})^2\right] \\ &= 4.24 \times 10^5 \text{ N/m}^2. \end{aligned}$$

Discussion

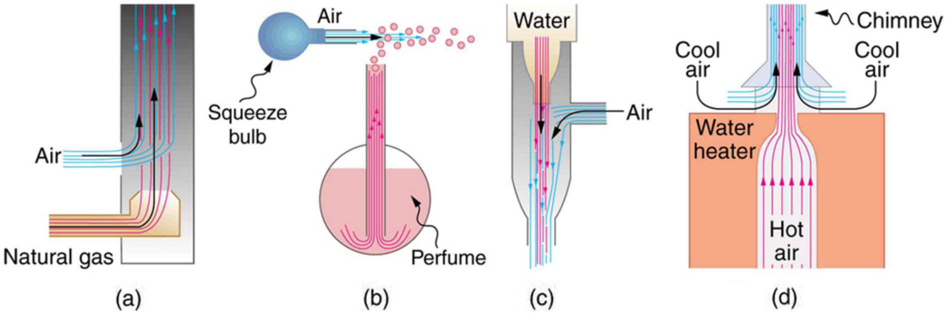
This absolute pressure in the hose is greater than in the nozzle, as expected since v is greater in the nozzle. The pressure P_2 in the nozzle must be atmospheric since it emerges into the atmosphere without other changes in conditions.

Applications of Bernoulli's Principle

There are a number of devices and situations in which fluid flows at a constant height and, thus, can be analyzed with Bernoulli's principle.

Entrainment

People have long put the Bernoulli principle to work by using reduced pressure in high-velocity fluids to move things about. With a higher pressure on the outside, the high-velocity fluid forces other fluids into the stream. This process is called *entrainment*. Entrainment devices have been in use since ancient times, particularly as pumps to raise water small heights, as in draining swamps, fields, or other low-lying areas. Some other devices that use the concept of entrainment are shown in [link].



Examples of entrainment devices that use increased fluid speed to create low pressures, which then entrain one fluid into another. (a) A Bunsen burner uses an adjustable gas nozzle, entraining air for proper combustion. (b) An atomizer uses a squeeze bulb to create a jet of air that entrains drops of perfume. Paint sprayers and carburetors use very similar techniques to move their respective liquids. (c) A common aspirator uses a high-speed stream of water to create a region of lower pressure. Aspirators may be used as suction pumps in dental and surgical situations or for draining a flooded basement or producing a reduced pressure in a vessel. (d) The chimney of a water heater is designed to entrain air into the pipe leading through the ceiling.

Wings and Sails

The airplane wing is a beautiful example of Bernoulli’s principle in action. [link](a) shows the characteristic shape of a wing. The wing is tilted upward at a small angle and the upper surface is longer, causing air to flow faster over it. The pressure on top of the wing is therefore reduced, creating a net upward force or lift. (Wings can also gain lift by pushing air downward, utilizing the conservation of momentum principle. The deflected air molecules result in an upward force on the wing — Newton’s third law.) Sails also have the characteristic shape of a wing. (See [link](b).) The pressure on the front side of the sail, P_{front}, is lower than the pressure on the back of the sail, P_{back}. This results in a forward force and even allows you to sail into the wind.

Note:

Making Connections: Take-Home Investigation with Two Strips of Paper

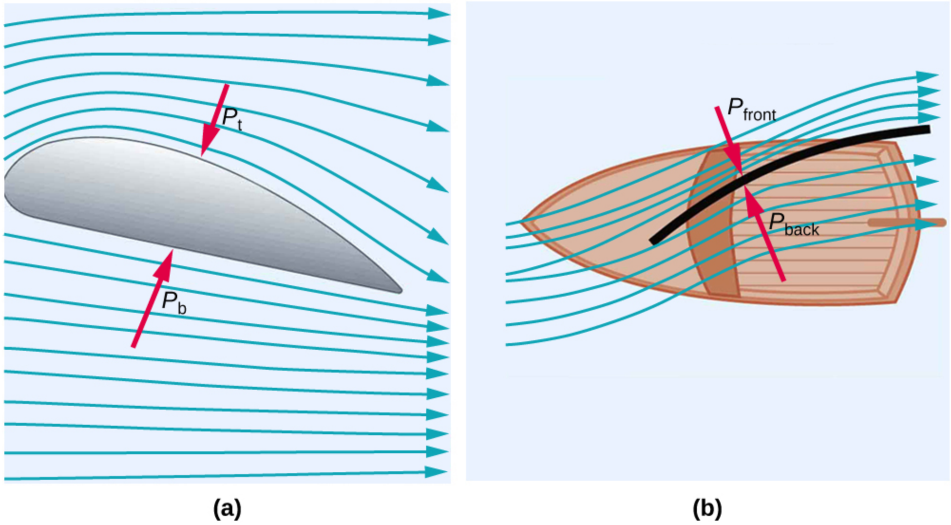
For a good illustration of Bernoulli’s principle, make two strips of paper, each about 15 cm long and 4 cm wide. Hold the small end of one strip up to your lips and let it drape over your finger. Blow across the paper. What happens? Now hold two strips of paper up to your lips, separated by your fingers. Blow between the strips. What happens?

Velocity measurement

[link] shows two devices that measure fluid velocity based on Bernoulli’s principle. The manometer in [link](a) is connected to two tubes that are small enough not to appreciably disturb the flow. The tube facing the oncoming fluid creates a dead spot having zero velocity ($v_1 = 0$) in front of it, while fluid passing the other tube has velocity v_2. This means that Bernoulli’s principle as stated in $P_1 + \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho v_2^2$ becomes

Equation:

$$P_1 = P_2 + \frac{1}{2}\rho v_2^2.$$



(a) The Bernoulli principle helps explain lift generated by a wing. (b) Sails use the same technique to generate part of their thrust.

Thus pressure P_2 over the second opening is reduced by $\frac{1}{2}\rho v_2^2$, and so the fluid in the manometer rises by h on the side connected to the second opening, where

Equation:

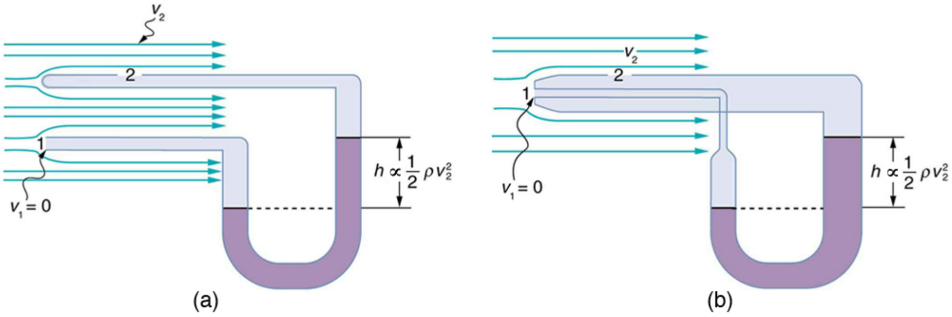
$$h \propto \frac{1}{2}\rho v_2^2.$$

(Recall that the symbol $\propto$ means "proportional to.") Solving for v_2, we see that

Equation:

$$v_2 \propto \sqrt{h}.$$

[link](b) shows a version of this device that is in common use for measuring various fluid velocities; such devices are frequently used as air speed indicators in aircraft.


v_2
2
1
$v_1 = 0$
$h \propto \frac{1}{2}\rho v_2^2$
(a)
v_2
2
1
$v_1 = 0$
$h \propto \frac{1}{2}\rho v_2^2$
(b)

Measurement of fluid speed based on Bernoulli's principle. (a) A manometer is connected to two tubes that are close together and small enough not to disturb the flow. Tube 1 is open at the end facing the flow. A dead spot having zero speed is created there. Tube 2 has an opening on the side, and so the fluid has a speed v across the opening; thus, pressure there drops. The difference in pressure at the manometer is $\frac{1}{2}\rho v_2^2$, and so h is proportional to $\frac{1}{2}\rho v_2^2$. (b) This type of velocity measuring device is a Prandtl tube, also known as a pitot tube.

Summary

- Bernoulli's equation states that the sum on each side of the following equation is constant, or the same at any two points in an incompressible frictionless fluid:
  Equation:

  $$P_1 + \frac{1}{2}\rho v_1^2 + \rho\, g h_1 = P_2 + \frac{1}{2}\rho v_2^2 + \rho \mathrm{gh}_2.$$

- Bernoulli’s principle is Bernoulli’s equation applied to situations in which depth is constant. The terms involving depth (or height h) subtract out, yielding
  Equation:
  $$P_1 + \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho v_2^2.$$
- Bernoulli’s principle has many applications, including entrainment, wings and sails, and velocity measurement.

Conceptual Questions

Exercise:

Problem:

You can squirt water a considerably greater distance by placing your thumb over the end of a garden hose and then releasing, than by leaving it completely uncovered. Explain how this works.

Exercise:

Problem:

Water is shot nearly vertically upward in a decorative fountain and the stream is observed to broaden as it rises. Conversely, a stream of water falling straight down from a faucet narrows. Explain why, and discuss whether surface tension enhances or reduces the effect in each case.

Exercise:

Problem:

Look back to [link]. Answer the following two questions. Why is P_{o} less than atmospheric? Why is P_{o} greater than P_{i}?

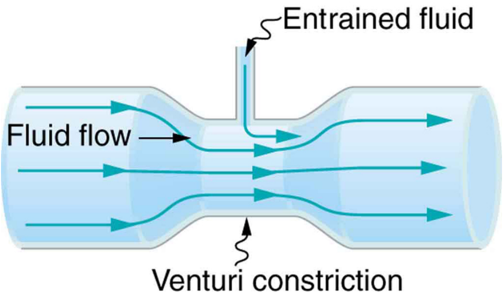
Exercise:

Problem: Give an example of entrainment not mentioned in the text.

Exercise:

Problem:

Many entrainment devices have a constriction, called a Venturi, such as shown in [link]. How does this bolster entrainment?



A tube with a narrow segment designed to enhance entrainment is called a Venturi. These are very commonly used in carburetors and aspirators.

Exercise:

Problem:

Some chimney pipes have a T-shape, with a crosspiece on top that helps draw up gases whenever there is even a slight breeze. Explain how this works in terms of Bernoulli’s principle.

Exercise:

Problem:

Is there a limit to the height to which an entrainment device can raise a fluid? Explain your answer.

Exercise:

Problem:

Why is it preferable for airplanes to take off into the wind rather than with the wind?

Exercise:

Problem:

Roofs are sometimes pushed off vertically during a tropical cyclone, and buildings sometimes explode outward when hit by a tornado. Use Bernoulli's principle to explain these phenomena.

Exercise:

Problem: Why does a sailboat need a keel?

Exercise:

Problem:

It is dangerous to stand close to railroad tracks when a rapidly moving commuter train passes. Explain why atmospheric pressure would push you toward the moving train.

Exercise:

Problem:

Water pressure inside a hose nozzle can be less than atmospheric pressure due to the Bernoulli effect. Explain in terms of energy how the water can emerge from the nozzle against the opposing atmospheric pressure.

Exercise:

Problem:

A perfume bottle or atomizer sprays a fluid that is in the bottle. ([link].) How does the fluid rise up in the vertical tube in the bottle?

Atomizer: perfume bottle with tube to carry perfume up through the bottle. (credit: Antonia Foy, Flickr)

Exercise:

Problem:

If you lower the window on a car while moving, an empty plastic bag can sometimes fly out the window. Why does this happen?

Problems & Exercises

Exercise:

Problem: Verify that pressure has units of energy per unit volume.

Solution:

$$\begin{aligned} P &= \frac{\text{Force}}{\text{Area}}, \\ (P)_{\text{units}} &= \text{N/m}^2 = \text{N}\cdot\text{m/m}^3 = \text{J/m}^3 \\ &= \text{energy/volume} \end{aligned}$$

Exercise:

Problem:

Suppose you have a wind speed gauge like the pitot tube shown in [link](b). By what factor must wind speed increase to double the value of h in the manometer? Is this independent of the moving fluid and the fluid in the manometer?

Exercise:

Problem:

If the pressure reading of your pitot tube is 15.0 mm Hg at a speed of 200 km/h, what will it be at 700 km/h at the same altitude?

Solution:

184 mm Hg

Exercise:

Problem:

Calculate the maximum height to which water could be squirted with the hose in [link] example if it: (a) Emerges from the nozzle. (b) Emerges with the nozzle removed, assuming the same flow rate.

Exercise:

Problem:

Every few years, winds in Boulder, Colorado, attain sustained speeds of 45.0 m/s (about 100 mi/h) when the jet stream descends during early spring. Approximately what is the force due to the Bernoulli effect on a roof having an area of $220\ \text{m}^2$? Typical air density in Boulder is $1.14\ \text{kg/m}^3$, and the corresponding atmospheric pressure is $8.89\times10^4\ \text{N/m}^2$. (Bernoulli's principle as stated in the text assumes laminar flow. Using the principle here produces only an approximate result, because there is significant turbulence.)

Solution:

$2.54\times10^5\ \text{N}$

Exercise:

Problem:

(a) Calculate the approximate force on a square meter of sail, given the horizontal velocity of the wind is 6.00 m/s parallel to its front surface and 3.50 m/s along its back surface. Take the density of air to be $1.29\ \text{kg/m}^3$. (The calculation, based on Bernoulli's principle, is approximate due to the effects of turbulence.) (b) Discuss whether this force is great enough to be effective for propelling a sailboat.

Exercise:

Problem:

(a) What is the pressure drop due to the Bernoulli effect as water goes into a 3.00-cm-diameter nozzle from a 9.00-cm-diameter fire hose while carrying a flow of 40.0 L/s? (b) To what maximum height above the nozzle can this water rise? (The actual height will be significantly smaller due to air resistance.)

Solution:

(a) $1.58\times10^6\ \text{N/m}^2$

(b) 163 m

Exercise:

Problem:

(a) Using Bernoulli’s equation, show that the measured fluid speed v for a pitot tube, like the one in [link](b), is given by

Equation:

$$v = \left(\frac{2\rho\prime gh}{\rho} \right)^{1/2},$$

where h is the height of the manometer fluid, $\rho\prime$ is the density of the manometer fluid, ρ is the density of the moving fluid, and g is the acceleration due to gravity. (Note that v is indeed proportional to the square root of h, as stated in the text.) (b) Calculate v for moving air if a mercury manometer’s h is 0.200 m.

Glossary

Bernoulli’s equation
: the equation resulting from applying conservation of energy to an incompressible frictionless fluid: $P + 1/2pv^2 + pgh = \text{constant}$, through the fluid

Bernoulli’s principle
: Bernoulli’s equation applied at constant depth: $P_1 + 1/2pv_1^2 = P_2 + 1/2pv_2^2$

The Most General Applications of Bernoulli’s Equation

- Calculate using Torricelli’s theorem.
- Calculate power in fluid flow.

Torricelli’s Theorem

[link] shows water gushing from a large tube through a dam. What is its speed as it emerges? Interestingly, if resistance is negligible, the speed is just what it would be if the water fell a distance h from the surface of the reservoir; the water’s speed is independent of the size of the opening. Let us check this out. Bernoulli’s equation must be used since the depth is not constant. We consider water flowing from the surface (point 1) to the tube’s outlet (point 2). Bernoulli’s equation as stated in previously is
Equation:

$$P_1 + \frac{1}{2}\rho v_1^2 + \rho g h_1 = P_2 + \frac{1}{2}\rho v_2^2 + \rho g h_2.$$

Both P_1 and P_2 equal atmospheric pressure (P_1 is atmospheric pressure because it is the pressure at the top of the reservoir. P_2 must be atmospheric pressure, since the emerging water is surrounded by the atmosphere and cannot have a pressure different from atmospheric pressure.) and subtract out of the equation, leaving
Equation:

$$\frac{1}{2}\rho v_1^2 + \rho g h_1 = \frac{1}{2}\rho v_2^2 + \rho g h_2.$$

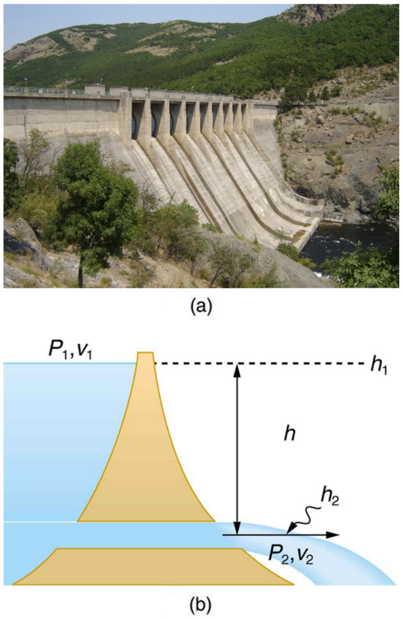
Solving this equation for v_2^2, noting that the density ρ cancels (because the fluid is incompressible), yields
Equation:

$$v_2^2 = v_1^2 + 2g(h_1 - h_2).$$

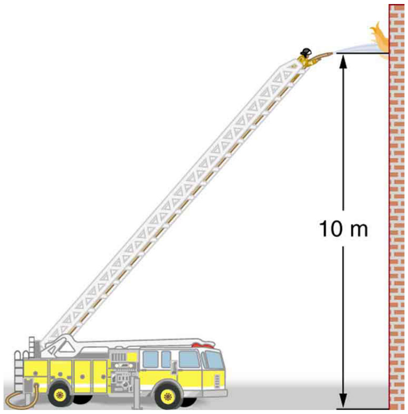
We let $h = h_1 - h_2$; the equation then becomes
Equation:

$$v_2^2 = v_1^2 + 2\text{gh}$$

where h is the height dropped by the water. This is simply a kinematic equation for any object falling a distance h with negligible resistance. In fluids, this last equation is called *Torricelli’s theorem*. Note that the result is independent of the velocity’s direction, just as we found when applying conservation of energy to falling objects.



(a) Water gushes from the base of the Studen Kladenetz dam in Bulgaria. (credit: Kiril Kapustin; http://www.ImagesFromBulgaria.com) (b) In the absence of significant resistance, water flows from the reservoir with the same speed it would have if it fell the distance h without friction. This is an example of Torricelli's theorem.



Pressure in the nozzle of this fire hose is less than at ground level for two reasons: the water has to go uphill to get to the nozzle, and speed increases in the nozzle. In spite of its lowered pressure, the water can exert a large force on anything it strikes, by virtue of its kinetic energy. Pressure in the water stream becomes equal to atmospheric pressure once it emerges into the air.

All preceding applications of Bernoulli’s equation involved simplifying conditions, such as constant height or constant pressure. The next example is a more general application of Bernoulli’s equation in which pressure, velocity, and height all change. (See [link].)

Example:
Calculating Pressure: A Fire Hose Nozzle
Fire hoses used in major structure fires have inside diameters of 6.40 cm. Suppose such a hose carries a flow of 40.0 L/s starting at a gauge pressure of $1.62\times10^{6}\ \mathrm{N/m^2}$. The hose goes 10.0 m up a ladder to a nozzle having an inside diameter of 3.00 cm. Assuming negligible resistance, what is the pressure in the nozzle?
Strategy
Here we must use Bernoulli’s equation to solve for the pressure, since depth is not constant.
Solution
Bernoulli’s equation states
Equation:

$$P_1+\frac{1}{2}\rho v_1^2+\rho g h_1=P_2+\frac{1}{2}\rho v_2^2+\rho g h_2,$$

where the subscripts 1 and 2 refer to the initial conditions at ground level and the final conditions inside the nozzle, respectively. We must first find the speeds v_1 and v_2. Since $Q=A_1 v_1$, we get
Equation:

$$v_1=\frac{Q}{A_1}=\frac{40.0\times10^{-3}\ \mathrm{m^3/s}}{\pi(3.20\times10^{-2}\ \mathrm{m})^2}=12.4\ \mathrm{m/s}.$$

Similarly, we find
Equation:

$$v_2=56.6\ \mathrm{m/s}.$$

(This rather large speed is helpful in reaching the fire.) Now, taking h_1 to be zero, we solve Bernoulli’s equation for P_2:
Equation:

$$P_2=P_1+\frac{1}{2}\rho\left(v_1^2-v_2^2\right)-\rho g h_2.$$

Substituting known values yields
Equation:

$$P_2=1.62\times10^{6}\ \mathrm{N/m^2}+\frac{1}{2}(1000\ \mathrm{kg/m^3})\big[(12.4\ \mathrm{m/s})^2-(56.6\ \mathrm{m/s})^2\big]-(1000\ \mathrm{kg/m^3})(9.80\ \mathrm{m/s^2})(10.0\ \mathrm{m}$$

Discussion

This value is a gauge pressure, since the initial pressure was given as a gauge pressure. Thus the nozzle pressure equals atmospheric pressure, as it must because the water exits into the atmosphere without changes in its conditions.

Power in Fluid Flow

Power is the *rate* at which work is done or energy in any form is used or supplied. To see the relationship of power to fluid flow, consider Bernoulli's equation:
Equation:

$$P+\frac{1}{2}\rho v^2+\rho \mathrm{gh}=\text{constant}.$$

All three terms have units of energy per unit volume, as discussed in the previous section. Now, considering units, if we multiply energy per unit volume by flow rate (volume per unit time), we get units of power. That is, $(E/V)(V/t)=E/t$. This means that if we multiply Bernoulli's equation by flow rate Q, we get power. In equation form, this is
Equation:

$$\left(P+\frac{1}{2}\rho v^2+\rho \mathrm{gh}\right)Q=\text{power}.$$

Each term has a clear physical meaning. For example, PQ is the power supplied to a fluid, perhaps by a pump, to give it its pressure P. Similarly, $\frac{1}{2}\rho v^2 Q$ is the power supplied to a fluid to give it its kinetic energy. And $\rho \mathrm{ghQ}$ is the power going to gravitational potential energy.

Note:
Making Connections: Power
Power is defined as the rate of energy transferred, or E/t. Fluid flow involves several types of power. Each type of power is identified with a specific type of energy being expended or changed in form.

Example:
Calculating Power in a Moving Fluid
Suppose the fire hose in the previous example is fed by a pump that receives water through a hose with a 6.40-cm diameter coming from a hydrant with a pressure of $0.700\times 10^6\ \mathrm{N/m^2}$. What power does the pump supply to the water?
Strategy
Here we must consider energy forms as well as how they relate to fluid flow. Since the input and output hoses have the same diameters and are at the same height, the pump does not change the speed of the water nor its height, and so the water's kinetic energy and gravitational potential energy are unchanged. That means the pump only supplies power to increase water pressure by $0.92\times 10^6\ \mathrm{N/m^2}$ (from $0.700\times 10^6\ \mathrm{N/m^2}$ to $1.62\times 10^6\ \mathrm{N/m^2}$).
Solution
As discussed above, the power associated with pressure is
Equation:

$$\begin{array}{rcl}\text{power} & = & \text{PQ} \\ & = & \left(0.920 \times 10^{6}\ \text{N/m}^{2}\right)\left(40.0 \times 10^{-3}\ \text{m}^{3}\text{/s}\right). \\ & = & 3.68 \times 10^{4}\ \text{W} = 36.8\ \text{kW}\end{array}$$

Discussion

Such a substantial amount of power requires a large pump, such as is found on some fire trucks. (This kilowatt value converts to about 50 hp.) The pump in this example increases only the water’s pressure. If a pump—such as the heart—directly increases velocity and height as well as pressure, we would have to calculate all three terms to find the power it supplies.

Summary

- Power in fluid flow is given by the equation $\left(P_1 + \frac{1}{2}\rho v^2 + \rho \text{gh}\right)Q = \text{power}$, where the first term is power associated with pressure, the second is power associated with velocity, and the third is power associated with height.

Conceptual Questions

Exercise:

Problem:

Based on Bernoulli’s equation, what are three forms of energy in a fluid? (Note that these forms are conservative, unlike heat transfer and other dissipative forms not included in Bernoulli’s equation.)

Exercise:

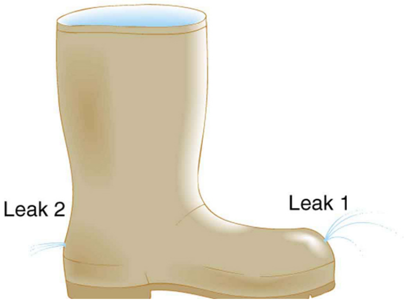
Problem:

Water that has emerged from a hose into the atmosphere has a gauge pressure of zero. Why? When you put your hand in front of the emerging stream you feel a force, yet the water’s gauge pressure is zero. Explain where the force comes from in terms of energy.

Exercise:

Problem:

The old rubber boot shown in [link] has two leaks. To what maximum height can the water squirt from Leak 1? How does the velocity of water emerging from Leak 2 differ from that of leak 1? Explain your responses in terms of energy.



Water emerges from two leaks in an old boot.

Exercise:

Problem:

Water pressure inside a hose nozzle can be less than atmospheric pressure due to the Bernoulli effect. Explain in terms of energy how the water can emerge from the nozzle against the opposing atmospheric pressure.

Problems & Exercises

Exercise:

Problem:

Hoover Dam on the Colorado River is the highest dam in the United States at 221 m, with an output of 1300 MW. The dam generates electricity with water taken from a depth of 150 m and an average flow rate of $650\ \text{m}^3/\text{s}$. (a) Calculate the power in this flow. (b) What is the ratio of this power to the facility's average of 680 MW?

Solution:

(a) $9.56 \times 10^8\ \text{W}$

(b) 1.4

Exercise:

Problem:

A frequently quoted rule of thumb in aircraft design is that wings should produce about 1000 N of lift per square meter of wing. (The fact that a wing has a top and bottom surface does not double its area.) (a) At takeoff, an aircraft travels at 60.0 m/s, so that the air speed relative to the bottom of the wing is 60.0 m/s. Given the sea level density of air to be $1.29\ \text{kg/m}^3$, how fast must it move over the upper surface to create the ideal lift? (b) How fast must air move over the upper surface at a cruising speed of 245 m/s and at an altitude where air density is one-fourth that at sea level? (Note that this is not all of the aircraft's lift—some comes from the body of the plane, some from engine thrust, and so on. Furthermore, Bernoulli's principle gives an approximate answer because flow over the wing creates turbulence.)

Exercise:

Problem:

The left ventricle of a resting adult's heart pumps blood at a flow rate of $83.0\ \text{cm}^3/\text{s}$, increasing its pressure by 110 mm Hg, its speed from zero to 30.0 cm/s, and its height by 5.00 cm. (All numbers are averaged over the entire heartbeat.) Calculate the total power output of the left ventricle. Note that most of the power is used to increase blood pressure.

Solution:

1.26 W

Exercise:

Problem:

A sump pump (used to drain water from the basement of houses built below the water table) is draining a flooded basement at the rate of 0.750 L/s, with an output pressure of $3.00 \times 10^5\ \text{N/m}^2$. (a) The water enters a hose with a 3.00-cm inside diameter and rises 2.50 m above the pump. What is its pressure at this point? (b) The hose goes over the foundation wall, losing 0.500 m in height, and widens to 4.00 cm in diameter. What is the pressure now? You may neglect frictional losses in both parts of the problem.

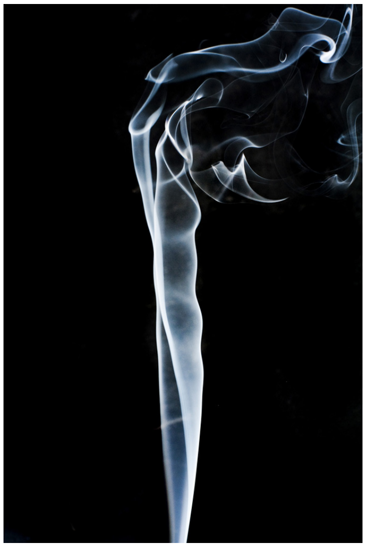
Viscosity and Laminar Flow; Poiseuille’s Law

- Define laminar flow and turbulent flow.
- Explain what viscosity is.
- Calculate flow and resistance with Poiseuille’s law.
- Explain how pressure drops due to resistance.

Laminar Flow and Viscosity

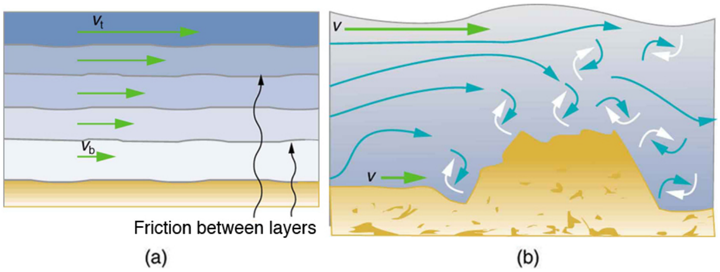
When you pour yourself a glass of juice, the liquid flows freely and quickly. But when you pour syrup on your pancakes, that liquid flows slowly and sticks to the pitcher. The difference is fluid friction, both within the fluid itself and between the fluid and its surroundings. We call this property of fluids *viscosity*. Juice has low viscosity, whereas syrup has high viscosity. In the previous sections we have considered ideal fluids with little or no viscosity. In this section, we will investigate what factors, including viscosity, affect the rate of fluid flow.

The precise definition of viscosity is based on *laminar*, or nonturbulent, flow. Before we can define viscosity, then, we need to define laminar flow and turbulent flow. [link] shows both types of flow. **Laminar** flow is characterized by the smooth flow of the fluid in layers that do not mix. Turbulent flow, or **turbulence**, is characterized by eddies and swirls that mix layers of fluid together.



Smoke rises smoothly for a while and then begins to form swirls and eddies. The smooth flow is called laminar flow, whereas the swirls and eddies typify turbulent flow. If you watch the smoke (being careful not to breathe on it), you will notice that it rises more rapidly when flowing smoothly than after it becomes turbulent, implying that turbulence poses more resistance to flow. (credit: Creativity103)

[link] shows schematically how laminar and turbulent flow differ. Layers flow without mixing when flow is laminar. When there is turbulence, the layers mix, and there are significant velocities in directions other than the overall direction of flow. The lines that are shown in many illustrations are the paths followed by small volumes of fluids. These are called *streamlines*. Streamlines are smooth and continuous when flow is laminar, but break up and mix when flow is turbulent. Turbulence has two main causes. First, any obstruction or sharp corner, such as in a faucet, creates turbulence by imparting velocities perpendicular to the flow. Second, high speeds cause turbulence. The drag both between adjacent layers of fluid and between the fluid and its surroundings forms swirls and eddies, if the speed is great enough. We shall concentrate on laminar flow for the remainder of this section, leaving certain aspects of turbulence for later sections.



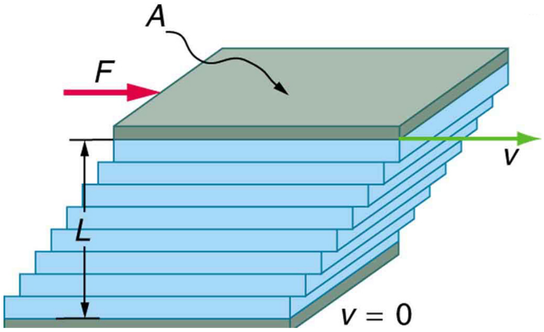
(a) Laminar flow occurs in layers without mixing. Notice that viscosity causes drag between layers as well as with the fixed surface. (b) An obstruction in the vessel produces turbulence. Turbulent flow mixes the fluid. There is more interaction, greater heating, and more resistance than in laminar flow.

Note:

Making Connections: Take-Home Experiment: Go Down to the River

Try dropping simultaneously two sticks into a flowing river, one near the edge of the river and one near the middle. Which one travels faster? Why?

[link] shows how viscosity is measured for a fluid. Two parallel plates have the specific fluid between them. The bottom plate is held fixed, while the top plate is moved to the right, dragging fluid with it. The layer (or lamina) of fluid in contact with either plate does not move relative to the plate, and so the top layer moves at v while the bottom layer remains at rest. Each successive layer from the top down exerts a force on the one below it, trying to drag it along, producing a continuous variation in speed from v to 0 as shown. Care is taken to insure that the flow is laminar; that is, the layers do not mix. The motion in [link] is like a continuous shearing motion. Fluids have zero shear strength, but the *rate* at which they are sheared is related to the same geometrical factors A and L as is shear deformation for solids.



The graphic shows laminar flow of fluid between two plates of area A. The bottom plate is fixed. When the top plate is pushed to the right, it drags the fluid along with it.

A force F is required to keep the top plate in [link] moving at a constant velocity v, and experiments have shown that this force depends on four factors. First, F is directly proportional to v (until the speed is so high that turbulence occurs—then a much larger force is needed, and it has a more complicated dependence on v). Second, F is proportional to the area A of the plate. This relationship seems reasonable, since A is directly proportional to the amount of fluid being moved. Third, F is inversely proportional to the distance between the plates L. This relationship is also reasonable; L is like a lever arm, and the greater the lever arm, the less force that is needed. Fourth, F is directly proportional to *the coefficient of viscosity*, η. The greater the viscosity, the greater the force required. These dependencies are combined into the equation

Equation:

$$F=\eta\frac{\text{vA}}{L},$$

which gives us a working definition of fluid **viscosity** η. Solving for η gives

Equation:

$$\eta =\frac{\text{FL}}{\text{vA}},$$

which defines viscosity in terms of how it is measured. The SI unit of viscosity is $\text{N}\cdot\text{m/}\left[\left(\text{m/s}\right)\text{m}^{2}\right]=\left(\text{N/m}^{2}\right)\text{s or Pa}\cdot\text{s}$. [link] lists the coefficients of viscosity for various fluids.

Viscosity varies from one fluid to another by several orders of magnitude. As you might expect, the viscosities of gases are much less than those of liquids, and these viscosities are often temperature dependent. The viscosity of blood can be reduced by aspirin consumption, allowing it to flow more easily around the body. (When used over the long term in low doses, aspirin can help prevent heart attacks, and reduce the risk of blood clotting.)

Laminar Flow Confined to Tubes—Poiseuille’s Law

What causes flow? The answer, not surprisingly, is pressure difference. In fact, there is a very simple relationship between horizontal flow and pressure. Flow rate Q is in the direction from high to low pressure. The greater the pressure differential between two points, the greater the flow rate. This relationship can be stated as
Equation:

$$Q=\frac{P_{2}-P_{1}}{R},$$

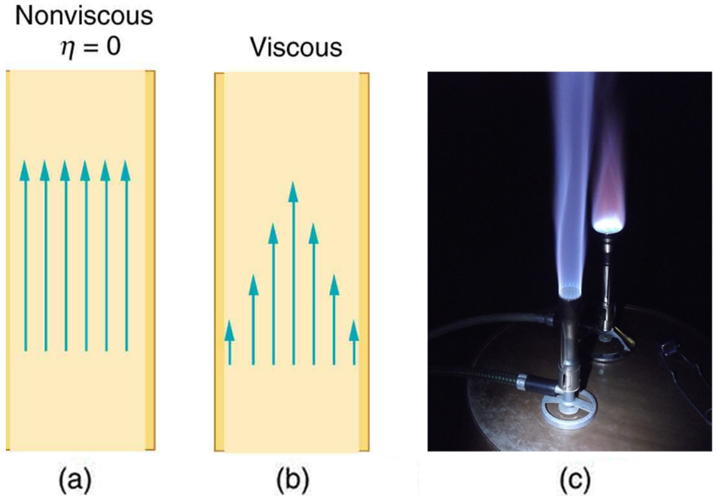
where P_{1} and P_{2} are the pressures at two points, such as at either end of a tube, and R is the resistance to flow. The resistance R includes everything, except pressure, that affects flow rate. For example, R is greater for a long tube than for a short one. The greater the viscosity of a fluid, the greater the value of R. Turbulence greatly increases R, whereas increasing the diameter of a tube decreases R.

If viscosity is zero, the fluid is frictionless and the resistance to flow is also zero. Comparing frictionless flow in a tube to viscous flow, as in [link], we see that for a viscous fluid, speed is greatest at midstream because of drag at the boundaries. We can see the effect of viscosity in a Bunsen burner flame, even though the viscosity of natural gas is small.

The resistance R to laminar flow of an incompressible fluid having viscosity η through a horizontal tube of uniform radius r and length l, such as the one in [link], is given by
Equation:

$$R=\frac{8\eta l}{\pi r^{4}}.$$

This equation is called **Poiseuille's law for resistance** after the French scientist J. L. Poiseuille (1799–1869), who derived it in an attempt to understand the flow of blood, an often turbulent fluid.



(a) If fluid flow in a tube has negligible resistance, the speed is the same all across the tube. (b) When a viscous fluid flows through a tube, its speed at the walls is zero, increasing steadily to its maximum at the center of the tube. (c) The shape of the Bunsen burner flame is due to the velocity profile across the tube. (credit: Jason Woodhead)

Let us examine Poiseuille's expression for R to see if it makes good intuitive sense. We see that resistance is directly proportional to both fluid viscosity η and the length l of a tube. After all, both of these directly affect the amount of friction encountered—the greater either is, the greater the resistance and the smaller the flow. The radius r of a tube affects the resistance, which again makes sense, because the greater the radius, the greater the flow (all other factors remaining the same). But it is surprising that r is raised to the *fourth* power in Poiseuille's law. This exponent means that any change in the radius of a tube has a very large effect on resistance. For example, doubling the radius of a tube decreases resistance by a factor of $2^4 = 16$.

Taken together, $Q = \frac{P_2 - P_1}{R}$ and $R = \frac{8\eta l}{\pi r^4}$ give the following expression for flow rate:
Equation:

$$Q = \frac{(P_2 - P_1)\pi r^4}{8\eta l}.$$

This equation describes laminar flow through a tube. It is sometimes called Poiseuille’s law for laminar flow, or simply **Poiseuille’s law**.

Example:
Using Flow Rate: Plaque Deposits Reduce Blood Flow
Suppose the flow rate of blood in a coronary artery has been reduced to half its normal value by plaque deposits. By what factor has the radius of the artery been reduced, assuming no turbulence occurs?
Strategy
Assuming laminar flow, Poiseuille’s law states that
Equation:

$$Q = \frac{(P_2 - P_1)\pi r^4}{8\eta l}.$$

We need to compare the artery radius before and after the flow rate reduction.
Solution
With a constant pressure difference assumed and the same length and viscosity, along the artery we have
Equation:

$$\frac{Q_1}{r_1^4} = \frac{Q_2}{r_2^4}.$$

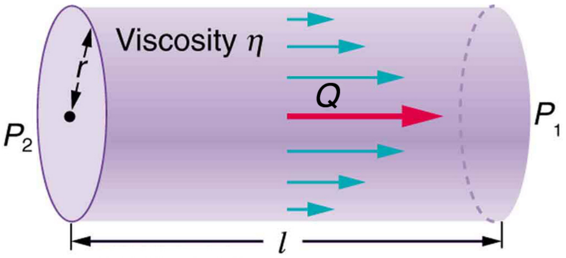
So, given that $Q_2 = 0.5Q_1$, we find that $r_2^4 = 0.5r_1^4$.
Therefore, $r_2 = (0.5)^{0.25} r_1 = 0.841 r_1$, a decrease in the artery radius of 16%.
Discussion
This decrease in radius is surprisingly small for this situation. To restore the blood flow in spite of this buildup would require an increase in the pressure difference $(P_2 - P_1)$ of a factor of two, with subsequent strain on the heart.

Fluid	**Temperature (ºC)**	**Viscosity** η (mPa·s)
Gases		
Air	0	0.0171
	20	0.0181
	40	0.0190
	100	0.0218
Ammonia	20	0.00974
Carbon dioxide	20	0.0147
Helium	20	0.0196
Hydrogen	0	0.0090
Mercury	20	0.0450
Oxygen	20	0.0203
Steam	100	0.0130
Liquids		
Water	0	1.792
	20	1.002
	37	0.6947
	40	0.653
	100	0.282
Whole blood[footnote]	20	3.015

Fluid	**Temperature (ºC)**	**Viscosity** η (mPa·s)
The ratios of the viscosities of blood to water are nearly constant between 0°C and 37°C.	37	2.084
Blood plasma[footnote] See note on Whole Blood.	20	1.810
	37	1.257
Ethyl alcohol	20	1.20
Methanol	20	0.584
Oil (heavy machine)	20	660
Oil (motor, SAE 10)	30	200
Oil (olive)	20	138
Glycerin	20	1500
Honey	20	2000–10000
Maple Syrup	20	2000–3000
Milk	20	3.0
Oil (Corn)	20	65

Coefficients of Viscosity of Various Fluids

The circulatory system provides many examples of Poiseuille’s law in action—with blood flow regulated by changes in vessel size and blood pressure. Blood vessels are not rigid but elastic. Adjustments to blood flow are primarily made by varying the size of the vessels, since the resistance is so sensitive to the radius. During vigorous exercise, blood vessels are selectively dilated to important muscles and organs and blood pressure increases. This creates both greater overall blood flow and increased flow to specific areas. Conversely, decreases in vessel radii, perhaps from plaques in the arteries, can greatly reduce blood flow. If a vessel's radius is reduced by only 5% (to 0.95 of its original value), the flow rate is reduced to about $(0.95)^4 = 0.81$ of its original value. A 19% decrease in flow is caused by a 5% decrease in radius. The body may compensate by increasing blood pressure by 19%, but this presents hazards to the heart and any vessel that has weakened walls. Another example comes from automobile engine oil. If you have a car with an oil pressure gauge, you may notice that oil pressure is high when the engine is cold. Motor oil has greater viscosity when cold than when warm, and so pressure must be greater to pump the same amount of cold oil.



Poiseuille's law applies to laminar flow of an incompressible fluid of viscosity η through a tube of length l and radius r. The direction of flow is from greater to lower pressure. Flow rate Q is directly proportional to the pressure difference $P_2 - P_1$, and inversely proportional to the length l of the tube and viscosity η of the fluid. Flow rate increases with r^4, the fourth power of the radius.

Example:
What Pressure Produces This Flow Rate?

An intravenous (IV) system is supplying saline solution to a patient at the rate of $0.120\ \text{cm}^3/\text{s}$ through a needle of radius 0.150 mm and length 2.50 cm. What pressure is needed at the entrance of the needle to cause this flow, assuming the viscosity of the saline solution to be the same as that of water? The gauge pressure of the blood in the patient’s vein is 8.00 mm Hg. (Assume that the temperature is 20ºC .)

Strategy

Assuming laminar flow, Poiseuille’s law applies. This is given by

Equation:

$$Q=\frac{(P_2-P_1)\pi r^4}{8\eta l},$$

where P_2 is the pressure at the entrance of the needle and P_1 is the pressure in the vein. The only unknown is P_2.

Solution

Solving for P_2 yields

Equation:

$$P_2=\frac{8\eta l}{\pi r^4}Q+P_{1.}$$

P_1 is given as 8.00 mm Hg, which converts to $1.066\times10^3\ \text{N/m}^2$. Substituting this and the other known values yields

Equation:

$$\begin{aligned}P_2 &= \left[\frac{8(1.00\times10^{-3}\ \text{N}\cdot\text{s/m}^2)(2.50\times10^{-2}\ \text{m})}{\pi(0.150\times10^{-3}\ \text{m})^4}\right](1.20\times10^{-7}\ \text{m}^3/\text{s})+1.066\times10^3\ \text{N/m}^2\\ &= 1.62\times10^4\ \text{N/m}^2.\end{aligned}$$

Discussion

This pressure could be supplied by an IV bottle with the surface of the saline solution 1.61 m above the entrance to the needle (this is left for you to solve in this chapter’s Problems and Exercises), assuming that there is negligible pressure drop in the tubing leading to the needle.

Flow and Resistance as Causes of Pressure Drops

You may have noticed that water pressure in your home might be lower than normal on hot summer days when there is more use. This pressure drop occurs in the water main before it reaches your home. Let us consider flow through the water main as illustrated in [link]. We can understand why the pressure P_1 to the home drops during times of heavy use by rearranging

Equation:

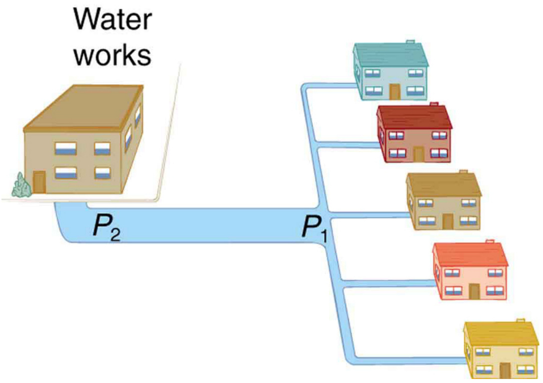
$$Q = \frac{P_2 - P_1}{R}$$

to

Equation:

$$P_2 - P_1 = RQ,$$

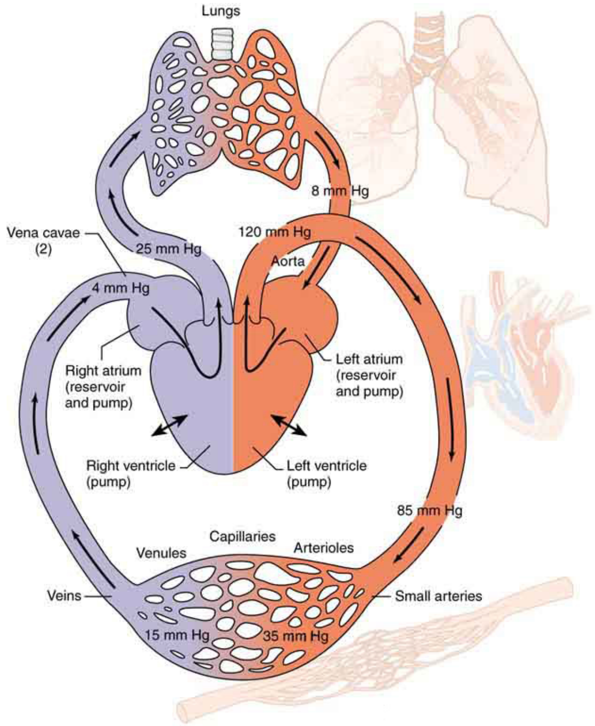
where, in this case, P_2 is the pressure at the water works and R is the resistance of the water main. During times of heavy use, the flow rate Q is large. This means that $P_2 - P_1$ must also be large. Thus P_1 must decrease. It is correct to think of flow and resistance as causing the pressure to drop from P_2 to P_1. $P_2 - P_1 = RQ$ is valid for both laminar and turbulent flows.



During times of heavy use, there is a significant pressure drop in a water main, and P_1 supplied to users is significantly less than P_2 created at the water works. If the flow is very small, then the pressure drop is negligible, and $P_2 \approx P_1$.

We can use $P_2 - P_1 = RQ$ to analyze pressure drops occurring in more complex systems in which the tube radius is not the same everywhere. Resistance will be much greater in narrow places, such as an obstructed coronary artery. For a given flow rate Q, the pressure drop will be greatest where the tube is most narrow. This is how water faucets control flow. Additionally, R is greatly increased by turbulence, and a constriction that creates turbulence greatly reduces the pressure downstream. Plaque in an artery reduces pressure and hence flow, both by its resistance and by the turbulence it creates.

[link] is a schematic of the human circulatory system, showing average blood pressures in its major parts for an adult at rest. Pressure created by the heart’s two pumps, the right and left ventricles, is reduced by the resistance of the blood vessels as the blood flows through them. The left ventricle increases arterial blood pressure that drives the flow of blood through all parts of the body except the lungs. The right ventricle receives the lower pressure blood from two major veins and pumps it through the lungs for gas exchange with atmospheric gases – the disposal of carbon dioxide from the blood and the replenishment of oxygen. Only one major organ is shown schematically, with typical branching of arteries to ever smaller vessels, the smallest of which are the capillaries, and rejoining of small veins into larger ones. Similar branching takes place in a variety of organs in the body, and the circulatory system has considerable flexibility in flow regulation to these organs by the dilation and constriction of the arteries leading to them and the capillaries within them. The sensitivity of flow to tube radius makes this flexibility possible over a large range of flow rates.


Lungs
8 mm Hg
Vena cavae
(2)
120 mm Hg
25 mm Hg
Aorta
4 mm Hg
Right atrium
(reservoir
and pump)
Left atrium
(reservoir
and pump)
Right ventricle
(pump)
Left ventricle
(pump)
85 mm Hg
Capillaries
Venules
Arterioles
Veins
Small arteries
15 mm Hg
35 mm Hg

Schematic of the circulatory system. Pressure difference is created by the two pumps in the heart and is reduced by resistance in the vessels. Branching of vessels into capillaries allows blood to reach individual cells and exchange substances, such as oxygen and waste products, with them. The system has an impressive ability to regulate flow to individual organs, accomplished largely by varying vessel diameters.

Each branching of larger vessels into smaller vessels increases the total cross-sectional area of the tubes through which the blood flows. For example, an artery with a cross section of $1\ \mathrm{cm}^2$ may branch into 20 smaller arteries, each with cross sections of $0.5\ \mathrm{cm}^2$, with a total of $10\ \mathrm{cm}^2$. In that manner, the resistance of the branchings is reduced so that pressure is not entirely lost. Moreover, because $Q = Av$ and A increases through branching, the average velocity of the blood in the smaller vessels is reduced. The blood velocity in the aorta ($\mathrm{diameter} = 1\ \mathrm{cm}$) is about 25 cm/s, while in the capillaries ($20 \mu\mathrm{m}$ in diameter) the velocity is about 1 mm/s. This reduced velocity allows the blood to exchange substances with the cells in the capillaries and alveoli in particular.

Section Summary

- Laminar flow is characterized by smooth flow of the fluid in layers that do not mix.
- Turbulence is characterized by eddies and swirls that mix layers of fluid together.
- Fluid viscosity η is due to friction within a fluid. Representative values are given in [link]. Viscosity has units of $(\mathrm{N/m^2})\mathrm{s}$ or $\mathrm{Pa \cdot s}$.
- Flow is proportional to pressure difference and inversely proportional to resistance:
  Equation:
  $$Q = \frac{P_2 - P_1}{R}.$$
- For laminar flow in a tube, Poiseuille's law for resistance states that
  Equation:
  $$R = \frac{8\eta l}{\pi r^4}.$$
- Poiseuille's law for flow in a tube is
  Equation:
  $$Q = \frac{(P_2 - P_1)\pi r^4}{8\eta l}.$$
- The pressure drop caused by flow and resistance is given by
  Equation:
  $$P_2 - P_1 = RQ.$$

Conceptual Questions

Exercise:

Problem:

Explain why the viscosity of a liquid decreases with temperature—that is, how might increased temperature reduce the effects of cohesive forces in a liquid? Also explain why the viscosity of a gas increases with temperature—that is, how does increased gas temperature create more collisions between atoms and molecules?

Exercise:

Problem:

When paddling a canoe upstream, it is wisest to travel as near to the shore as possible. When canoeing downstream, it may be best to stay near the middle. Explain why.

Exercise:

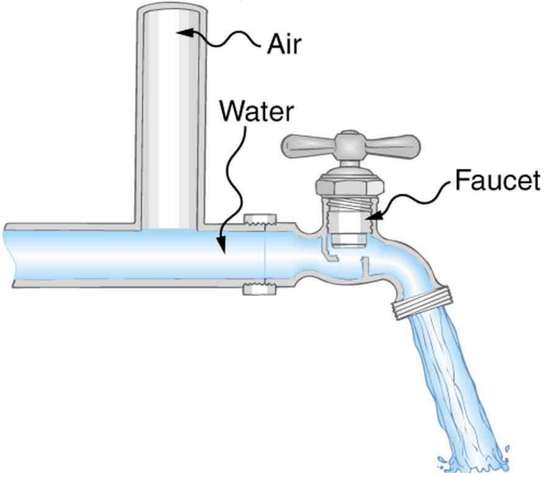
Problem:

Why does flow decrease in your shower when someone flushes the toilet?

Exercise:

Problem:

Plumbing usually includes air-filled tubes near water faucets, as shown in [link]. Explain why they are needed and how they work.



The vertical tube near the water tap remains full of air and serves a useful purpose.

Problems & Exercises

Exercise:

Problem:

(a) Calculate the retarding force due to the viscosity of the air layer between a cart and a level air track given the following information—air temperature is $20^{\text{o}}\ \text{C}$, the cart is moving at 0.400 m/s, its surface area is $2.50 \times 10^{-2}\ \text{m}^2$, and the thickness of the air layer is $6.00 \times 10^{-5}\ \text{m}$. (b) What is the ratio of this force to the weight of the 0.300-kg cart?

Solution:

(a) $3.02 \times 10^{-3}\ \text{N}$

(b) 1.03×10^{-3}

Exercise:

Problem:

What force is needed to pull one microscope slide over another at a speed of 1.00 cm/s, if there is a 0.500-mm-thick layer of $20^{\text{o}}\ \text{C}$ water between them and the contact area is $8.00\ \text{cm}^2$?

Exercise:

Problem:

A glucose solution being administered with an IV has a flow rate of $4.00\ \text{cm}^3/\text{min}$. What will the new flow rate be if the glucose is replaced by whole blood having the same density but a viscosity 2.50 times that of the glucose? All other factors remain constant.

Solution:

$1.60\ \text{cm}^3/\text{min}$

Exercise:

Problem:

The pressure drop along a length of artery is 100 Pa, the radius is 10 mm, and the flow is laminar. The average speed of the blood is 15 mm/s. (a) What is the net force on the blood in this section of artery? (b) What is the power expended maintaining the flow?

Exercise:

Problem:

A small artery has a length of $1.1 \times 10^{-3}\ \text{m}$ and a radius of $2.5 \times 10^{-5}\ \text{m}$. If the pressure drop across the artery is 1.3 kPa, what is the flow rate through the artery? (Assume that the temperature is $37^{\text{o}}\ \text{C}$.)

Solution:

$8.7 \times 10^{-11}\ \text{m}^3/\text{s}$

Exercise:

Problem:

Fluid originally flows through a tube at a rate of $100\ \text{cm}^3/\text{s}$. To illustrate the sensitivity of flow rate to various factors, calculate the new flow rate for the following changes with all other factors remaining the same as in the original conditions. (a) Pressure difference increases by a factor of 1.50. (b) A new fluid with 3.00 times greater viscosity is substituted. (c) The tube is replaced by one having 4.00 times the length. (d) Another tube is used with a radius 0.100 times the original. (e) Yet another tube is substituted with a radius 0.100 times the original and half the length, *and* the pressure difference is increased by a factor of 1.50.

Exercise:

Problem:

The arterioles (small arteries) leading to an organ, constrict in order to decrease flow to the organ. To shut down an organ, blood flow is reduced naturally to 1.00% of its original value. By what factor did the radii of the arterioles constrict? Penguins do this when they stand on ice to reduce the blood flow to their feet.

Solution:

0.316

Exercise:

Problem:

Angioplasty is a technique in which arteries partially blocked with plaque are dilated to increase blood flow. By what factor must the radius of an artery be increased in order to increase blood flow by a factor of 10?

Exercise:

Problem:

(a) Suppose a blood vessel’s radius is decreased to 90.0% of its original value by plaque deposits and the body compensates by increasing the pressure difference along the vessel to keep the flow rate constant. By what factor must the pressure difference increase? (b) If turbulence is created by the obstruction, what additional effect would it have on the flow rate?

Solution:

(a) 1.52

(b) Turbulence will decrease the flow rate of the blood, which would require an even larger increase in the pressure difference, leading to higher blood pressure.

Exercise:

Problem:

A spherical particle falling at a terminal speed in a liquid must have the gravitational force balanced by the drag force and the buoyant force. The buoyant force is equal to the weight of the displaced fluid, while the drag force is assumed to be given by Stokes Law, $F_s = 6\pi r\eta v$. Show that the terminal speed is given by

Equation:

$$v = \frac{2R^2 g}{9\eta}(\rho_{\text{s}} - \rho_1),$$

where R is the radius of the sphere, ρ_{s} is its density, and ρ_1 is the density of the fluid and η the coefficient of viscosity.

Exercise:

Problem:

Using the equation of the previous problem, find the viscosity of motor oil in which a steel ball of radius 0.8 mm falls with a terminal speed of 4.32 cm/s. The densities of the ball and the oil are 7.86 and 0.88 g/mL, respectively.

Solution:
Equation:

$$225\ \text{mPa}\cdot\text{s}$$

Exercise:

Problem:

A skydiver will reach a terminal velocity when the air drag equals their weight. For a skydiver with high speed and a large body, turbulence is a factor. The drag force then is approximately proportional to the square of the velocity. Taking the drag force to be $F_\text{D} = \frac{1}{2}\rho Av^2$ and setting this equal to the person's weight, find the terminal speed for a person falling "spread eagle." Find both a formula and a number for v_t, with assumptions as to size.

Exercise:

Problem:

A layer of oil 1.50 mm thick is placed between two microscope slides. Researchers find that a force of $5.50 \times 10^{-4}\ \text{N}$ is required to glide one over the other at a speed of 1.00 cm/s when their contact area is $6.00\ \text{cm}^2$. What is the oil's viscosity? What type of oil might it be?

Solution:
Equation:

$$0.138\ \text{Pa}\cdot\text{s},$$

or

Olive oil.

Exercise:

Problem:

(a) Verify that a 19.0% decrease in laminar flow through a tube is caused by a 5.00% decrease in radius, assuming that all other factors remain constant, as stated in the text. (b) What increase in flow is obtained from a 5.00% increase in radius, again assuming all other factors remain constant?

Exercise:

Problem:

[link] dealt with the flow of saline solution in an IV system. (a) Verify that a pressure of $1.62\times 10^{4}\ \text{N/m}^2$ is created at a depth of 1.61 m in a saline solution, assuming its density to be that of sea water. (b) Calculate the new flow rate if the height of the saline solution is decreased to 1.50 m. (c) At what height would the direction of flow be reversed? (This reversal can be a problem when patients stand up.)

Solution:

(a) $1.62\times 10^{4}\ \text{N/m}^2$

(b) $0.111\ \text{cm}^3\text{/s}$

(c)10.6 cm

Exercise:

Problem:

When physicians diagnose arterial blockages, they quote the reduction in flow rate. If the flow rate in an artery has been reduced to 10.0% of its normal value by a blood clot and the average pressure difference has increased by 20.0%, by what factor has the clot reduced the radius of the artery?

Exercise:

Problem:

During a marathon race, a runner’s blood flow increases to 10.0 times her resting rate. Her blood’s viscosity has dropped to 95.0% of its normal value, and the blood pressure difference across the circulatory system has increased by 50.0%. By what factor has the average radii of her blood vessels increased?

Solution:

1.59

Exercise:

Problem:

Water supplied to a house by a water main has a pressure of $3.00 \times 10^5 \text{ N/m}^2$ early on a summer day when neighborhood use is low. This pressure produces a flow of 20.0 L/min through a garden hose. Later in the day, pressure at the exit of the water main and entrance to the house drops, and a flow of only 8.00 L/min is obtained through the same hose. (a) What pressure is now being supplied to the house, assuming resistance is constant? (b) By what factor did the flow rate in the water main increase in order to cause this decrease in delivered pressure? The pressure at the entrance of the water main is $5.00 \times 10^5 \text{ N/m}^2$, and the original flow rate was 200 L/min. (c) How many more users are there, assuming each would consume 20.0 L/min in the morning?

Exercise:

Problem:

An oil gusher shoots crude oil 25.0 m into the air through a pipe with a 0.100-m diameter. Neglecting air resistance but not the resistance of the pipe, and assuming laminar flow, calculate the gauge pressure at the entrance of the 50.0-m-long vertical pipe. Take the density of the oil to be 900 kg/m^3 and its viscosity to be $1.00 \left(\text{N/m}^2\right) \cdot \text{s}$ (or $1.00 \text{ Pa} \cdot \text{s}$). Note that you must take into account the pressure due to the 50.0-m column of oil in the pipe.

Solution:

$2.95 \times 10^6 \text{ N/m}^2$(gauge pressure)

Exercise:

Problem:

Concrete is pumped from a cement mixer to the place it is being laid, instead of being carried in wheelbarrows. The flow rate is 200.0 L/min through a 50.0-m-long, 8.00-cm-diameter hose, and the pressure at the pump is $8.00 \times 10^6 \text{ N/m}^2$. (a) Calculate the resistance of the hose. (b) What is the viscosity of the concrete, assuming the flow is laminar? (c) How much power is being supplied, assuming the point of use is at the same level as the pump? You may neglect the power supplied to increase the concrete’s velocity.

Exercise:

Problem: Construct Your Own Problem

Consider a coronary artery constricted by arteriosclerosis. Construct a problem in which you calculate the amount by which the diameter of the artery is decreased, based on an assessment of the decrease in flow rate.

Exercise:

Problem:

Consider a river that spreads out in a delta region on its way to the sea. Construct a problem in which you calculate the average speed at which water moves in the delta region, based on the speed at which it was moving up river. Among the things to consider are the size and flow rate of the river before it spreads out and its size once it has spread out. You can construct the problem for the river spreading out into one large river or into multiple smaller rivers.

Glossary

laminar

a type of fluid flow in which layers do not mix

turbulence

fluid flow in which layers mix together via eddies and swirls

viscosity

the friction in a fluid, defined in terms of the friction between layers

Poiseuille’s law for resistance

the resistance to laminar flow of an incompressible fluid in a tube: $R = 8\eta l/\pi r^4$

Poiseuille’s law

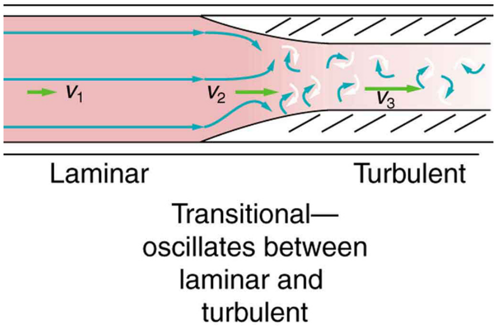
the rate of laminar flow of an incompressible fluid in a tube: $Q = (P_2 - P_1)\pi r^4/8\eta l$

The Onset of Turbulence

- Calculate Reynolds number.
- Use the Reynolds number for a system to determine whether it is laminar or turbulent.

Sometimes we can predict if flow will be laminar or turbulent. We know that flow in a very smooth tube or around a smooth, streamlined object will be laminar at low velocity. We also know that at high velocity, even flow in a smooth tube or around a smooth object will experience turbulence. In between, it is more difficult to predict. In fact, at intermediate velocities, flow may oscillate back and forth indefinitely between laminar and turbulent.

An occlusion, or narrowing, of an artery, such as shown in [link], is likely to cause turbulence because of the irregularity of the blockage, as well as the complexity of blood as a fluid. Turbulence in the circulatory system is noisy and can sometimes be detected with a stethoscope, such as when measuring diastolic pressure in the upper arm's partially collapsed brachial artery. These turbulent sounds, at the onset of blood flow when the cuff pressure becomes sufficiently small, are called *Korotkoff sounds*. Aneurysms, or ballooning of arteries, create significant turbulence and can sometimes be detected with a stethoscope. Heart murmurs, consistent with their name, are sounds produced by turbulent flow around damaged and insufficiently closed heart valves. Ultrasound can also be used to detect turbulence as a medical indicator in a process analogous to Doppler-shift radar used to detect storms.



Flow is laminar in the large part of this blood vessel and turbulent in the part narrowed by plaque, where velocity is high. In the transition region, the flow can oscillate chaotically between laminar and turbulent flow.

An indicator called the **Reynolds number** N_{R} can reveal whether flow is laminar or turbulent. For flow in a tube of uniform diameter, the Reynolds number is defined as

Equation:

$$N_{\text{R}} = \frac{2\rho \text{vr}}{\eta} (\text{flow in tube}),$$

where ρ is the fluid density, v its speed, η its viscosity, and r the tube radius. The Reynolds number is a unitless quantity. Experiments have revealed that N_{R} is related to the onset of turbulence. For N_{R} below about 2000, flow is laminar. For N_{R} above about 3000, flow is turbulent. For values of N_{R} between about 2000 and 3000, flow is unstable—that is, it can be laminar, but small obstructions and surface roughness can make it turbulent, and it may oscillate randomly between being laminar and turbulent. The blood flow through most of the body is a quiet, laminar flow. The exception is in the aorta, where the speed of the blood flow rises above a critical value of 35 m/s and becomes turbulent.

Example:

Is This Flow Laminar or Turbulent?

Calculate the Reynolds number for flow in the needle considered in Example 12.8 to verify the assumption that the flow is laminar. Assume that the density of the saline solution is 1025 kg/m^3.

Strategy

We have all of the information needed, except the fluid speed v, which can be calculated from $v = Q/A = 1.70 \text{ m/s}$ (verification of this is in this chapter's Problems and Exercises).

Solution

Entering the known values into $N_{\text{R}} = \frac{2\rho \text{vr}}{\eta}$ gives

Equation:

$$\begin{aligned} N_{\text{R}} &= \frac{2\rho\text{vr}}{\eta} \\ &= \frac{2(1025 \text{ kg/m}^3)(1.70 \text{ m/s})(0.150\times 10^{-3} \text{ m})}{1.00\times 10^{-3} \text{ N}\cdot\text{s/m}^2} \\ &= 523. \end{aligned}$$

Discussion

Since N_{R} is well below 2000, the flow should indeed be laminar.

Note:

Take-Home Experiment: Inhalation

Under the conditions of normal activity, an adult inhales about 1 L of air during each inhalation. With the aid of a watch, determine the time for one of your own inhalations by timing several breaths and dividing the total length by the number of breaths. Calculate the average flow rate Q of air traveling through the trachea during each inhalation.

The topic of chaos has become quite popular over the last few decades. A system is defined to be *chaotic* when its behavior is so sensitive to some factor that it is extremely difficult to predict. The field of *chaos* is the study of chaotic behavior. A good example of chaotic behavior is the flow of a fluid with a Reynolds number between 2000 and 3000. Whether or not the flow is turbulent is difficult, but not impossible, to predict—the difficulty lies in the extremely sensitive dependence on factors like roughness and obstructions on the nature of the flow. A tiny variation in one factor has an exaggerated (or nonlinear) effect on the flow. Phenomena as disparate as turbulence, the orbit of Pluto, and the onset of irregular heartbeats are chaotic and can be analyzed with similar techniques.

Section Summary

- The Reynolds number N_{R} can reveal whether flow is laminar or turbulent. It is
  Equation:

  $$N_{\text{R}}=\frac{2\rho \text{vr}}{\eta}.$$

- For N_{R} below about 2000, flow is laminar. For N_{R} above about 3000, flow is turbulent. For values of N_{R} between 2000 and 3000, it may be either or both.

Conceptual Questions

Exercise:

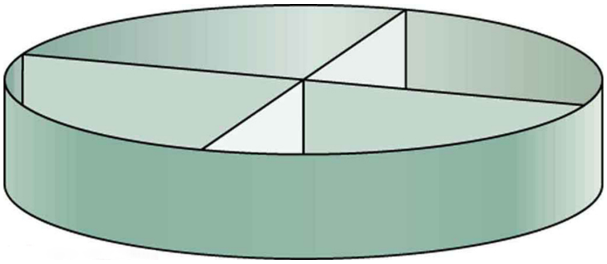
Problem:

Doppler ultrasound can be used to measure the speed of blood in the body. If there is a partial constriction of an artery, where would you expect blood speed to be greatest, at or nearby the constriction? What are the two distinct causes of higher resistance in the constriction?

Exercise:

Problem:

Sink drains often have a device such as that shown in [link] to help speed the flow of water. How does this work?



You will find devices such as this in many drains. They significantly increase flow rate.

Exercise:

Problem:

Some ceiling fans have decorative wicker reeds on their blades. Discuss whether these fans are as quiet and efficient as those with smooth blades.

Problems & Exercises

Exercise:

Problem:

Verify that the flow of oil is laminar (barely) for an oil gusher that shoots crude oil 25.0 m into the air through a pipe with a 0.100-m diameter. The vertical pipe is 50 m long. Take the density of the oil to be 900 kg/m^3 and its viscosity to be $1.00\ (\text{N/m}^2)\cdot\text{s}$ (or $1.00\ \text{Pa}\cdot\text{s}$).

Solution:

$N_{\text{R}} = 1.99 \times 10^2 < 2000$

Exercise:

Problem:

Show that the Reynolds number N_{R} is unitless by substituting units for all the quantities in its definition and cancelling.

Exercise:

Problem:

Calculate the Reynolds numbers for the flow of water through (a) a nozzle with a radius of 0.250 cm and (b) a garden hose with a radius of 0.900 cm, when the nozzle is attached to the hose. The flow rate through hose and nozzle is 0.500 L/s. Can the flow in either possibly be laminar?

Solution:

(a) nozzle: 1.27×10^{5} , not laminar

(b) hose: 3.51×10^{4}, not laminar.

Exercise:

Problem:

A fire hose has an inside diameter of 6.40 cm. Suppose such a hose carries a flow of 40.0 L/s starting at a gauge pressure of $1.62 \times 10^{6}\ \mathrm{N/m^{2}}$. The hose goes 10.0 m up a ladder to a nozzle having an inside diameter of 3.00 cm. Calculate the Reynolds numbers for flow in the fire hose and nozzle to show that the flow in each must be turbulent.

Exercise:

Problem:

Concrete is pumped from a cement mixer to the place it is being laid, instead of being carried in wheelbarrows. The flow rate is 200.0 L/min through a 50.0-m-long, 8.00-cm-diameter hose, and the pressure at the pump is $8.00 \times 10^6 \text{ N/m}^2$. Verify that the flow of concrete is laminar taking concrete's viscosity to be $48.0 \left(\text{N/m}^2\right) \cdot \text{s}$, and given its density is 2300 kg/m^3.

Solution:

2.54 << 2000, laminar.

Exercise:

Problem:

At what flow rate might turbulence begin to develop in a water main with a 0.200-m diameter? Assume a 20^o C temperature.

Exercise:

Problem:

What is the greatest average speed of blood flow at 37^o C in an artery of radius 2.00 mm if the flow is to remain laminar? What is the corresponding flow rate? Take the density of blood to be 1025 kg/m^3.

Solution:

1.02 m/s

$1.28 \times 10^{-2} \text{ L/s}$

Exercise:

Problem:

In Take-Home Experiment: Inhalation, we measured the average flow rate Q of air traveling through the trachea during each inhalation. Now calculate the average air speed in meters per second through your trachea during each inhalation. The radius of the trachea in adult humans is approximately 10^{-2} m. From the data above, calculate the Reynolds number for the air flow in the trachea during inhalation. Do you expect the air flow to be laminar or turbulent?

Exercise:

Problem:

Gasoline is piped underground from refineries to major users. The flow rate is $3.00 \times 10^{-2}\text{ m}^3/\text{s}$ (about 500 gal/min), the viscosity of gasoline is $1.00 \times 10^{-3}\ (\text{N/m}^2) \cdot \text{s}$, and its density is 680 kg/m^3. (a) What minimum diameter must the pipe have if the Reynolds number is to be less than 2000? (b) What pressure difference must be maintained along each kilometer of the pipe to maintain this flow rate?

Solution:

(a)$\geq 13.0\text{ m}$

(b) $2.68 \times 10^{-6}\text{ N/m}^2$

Exercise:

Problem:

Assuming that blood is an ideal fluid, calculate the critical flow rate at which turbulence is a certainty in the aorta. Take the diameter of the aorta to be 2.50 cm. (Turbulence will actually occur at lower average flow rates, because blood is not an ideal fluid. Furthermore, since blood flow pulses, turbulence may occur during only the high-velocity part of each heartbeat.)

Exercise:

Problem: **Unreasonable Results**

A fairly large garden hose has an internal radius of 0.600 cm and a length of 23.0 m. The nozzleless horizontal hose is attached to a faucet, and it delivers 50.0 L/s. (a) What water pressure is supplied by the faucet? (b) What is unreasonable about this pressure? (c) What is unreasonable about the premise? (d) What is the Reynolds number for the given flow? (Take the viscosity of water as $1.005 \times 10^{-3}\ \left(\mathrm{N/m^2}\right) \cdot \mathrm{s}$.)

Solution:

(a) 23.7 atm or $344\ \mathrm{lb/in^2}$

(b) The pressure is much too high.

(c) The assumed flow rate is very high for a garden hose.

(d) 5.27×10^6 > > 3000, turbulent, contrary to the assumption of laminar flow when using this equation.

Glossary

Reynolds number	a dimensionless parameter that can reveal whether a particular flow is laminar or turbulent

Motion of an Object in a Viscous Fluid

- Calculate the Reynolds number for an object moving through a fluid.
- Explain whether the Reynolds number indicates laminar or turbulent flow.
- Describe the conditions under which an object has a terminal speed.

A moving object in a viscous fluid is equivalent to a stationary object in a flowing fluid stream. (For example, when you ride a bicycle at 10 m/s in still air, you feel the air in your face exactly as if you were stationary in a 10-m/s wind.) Flow of the stationary fluid around a moving object may be laminar, turbulent, or a combination of the two. Just as with flow in tubes, it is possible to predict when a moving object creates turbulence. We use another form of the Reynolds number $N\prime_{\text{R}}$, defined for an object moving in a fluid to be

Equation:

$$N\prime_{\text{R}} = \frac{\rho \text{vL}}{\eta}(\text{object in fluid}),$$

where L is a characteristic length of the object (a sphere's diameter, for example), ρ the fluid density, η its viscosity, and v the object's speed in the fluid. If $N\prime_{\text{R}}$ is less than about 1, flow around the object can be laminar, particularly if the object has a smooth shape. The transition to turbulent flow occurs for $N\prime_{\text{R}}$ between 1 and about 10, depending on surface roughness and so on. Depending on the surface, there can be a *turbulent wake* behind the object with some laminar flow over its surface. For an $N\prime_{\text{R}}$ between 10 and 10^6, the flow may be either laminar or turbulent and may oscillate between the two. For $N\prime_{\text{R}}$ greater than about 10^6, the flow is entirely turbulent, even at the surface of the object. (See [link].) Laminar flow occurs mostly when the objects in the fluid are small, such as raindrops, pollen, and blood cells in plasma.

Example:
Does a Ball Have a Turbulent Wake?

Calculate the Reynolds number N'_{R} for a ball with a 7.40-cm diameter thrown at 40.0 m/s.

Strategy

We can use $N'_{\text{R}} = \frac{\rho \text{vL}}{\eta}$ to calculate N'_{R}, since all values in it are either given or can be found in tables of density and viscosity.

Solution

Substituting values into the equation for N'_{R} yields

Equation:

$$\begin{aligned} N'_R &= \frac{\rho \text{vL}}{\eta} = \frac{(1.29\ \text{kg/m}^3)(40.0\ \text{m/s})(0.0740\ \text{m})}{1.81 \times 10^{-5}\ 1.00\ \text{Pa}\cdot\text{s}} \\ &= 2.11 \times 10^5. \end{aligned}$$

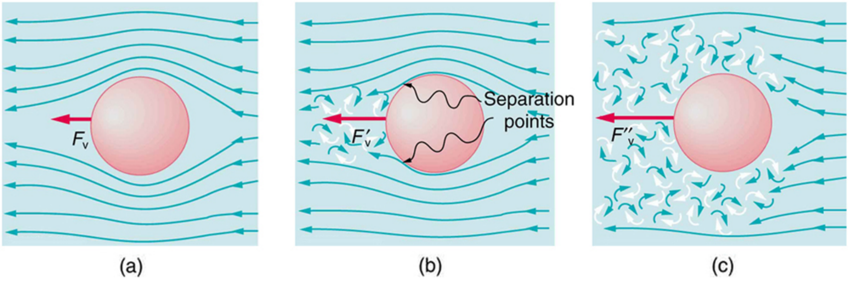
Discussion

This value is sufficiently high to imply a turbulent wake. Most large objects, such as airplanes and sailboats, create significant turbulence as they move. As noted before, the Bernoulli principle gives only qualitatively-correct results in such situations.

One of the consequences of viscosity is a resistance force called **viscous drag** F_{V} that is exerted on a moving object. This force typically depends on the object’s speed (in contrast with simple friction). Experiments have shown that for laminar flow (N'_{R} less than about one) viscous drag is proportional to speed, whereas for N'_{R} between about 10 and 10^6, viscous drag is proportional to speed squared. (This relationship is a strong dependence and is pertinent to bicycle racing, where even a small headwind causes significantly increased drag on the racer. Cyclists take turns being the leader in the pack for this reason.) For N'_{R} greater than 10^6, drag increases dramatically and behaves with greater complexity. For laminar flow around a sphere, F_{V} is proportional to fluid viscosity η, the object’s characteristic size L, and its speed v. All of which makes sense—the more viscous the fluid and the larger the object, the more drag we expect. Recall Stoke’s law $F_{\text{S}} = 6\pi r\eta v$. For the special case of a small sphere of radius R moving slowly in a fluid of viscosity η, the drag force F_{S} is given by

Equation:

$$F_{\mathrm{S}} = 6\pi R\eta v.$$



(a) Motion of this sphere to the right is equivalent to fluid flow to the left. Here the flow is laminar with N'_{R} less than 1. There is a force, called viscous drag F_{V}, to the left on the ball due to the fluid's viscosity. (b) At a higher speed, the flow becomes partially turbulent, creating a wake starting where the flow lines separate from the surface. Pressure in the wake is less than in front of the sphere, because fluid speed is less, creating a net force to the left F'_{V} that is significantly greater than for laminar flow. Here N'_{R} is greater than 10. (c) At much higher speeds, where N'_{R} is greater than 10^6, flow becomes turbulent everywhere on the surface and behind the sphere. Drag increases dramatically.

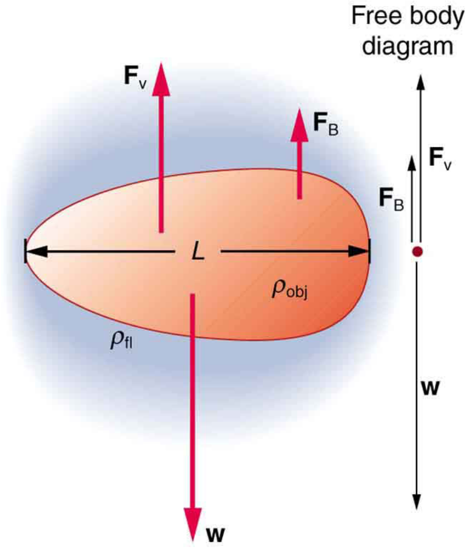
An interesting consequence of the increase in F_{V} with speed is that an object falling through a fluid will not continue to accelerate indefinitely (as it would if we neglect air resistance, for example). Instead, viscous drag increases, slowing acceleration, until a critical speed, called the **terminal speed**, is reached and the acceleration of the object becomes zero. Once this happens, the object continues to fall at constant speed (the terminal speed). This is the case for particles of sand falling in the ocean, cells falling in a centrifuge, and sky divers falling through the air. [link] shows some of the factors that affect terminal speed. There is a viscous drag on the object that depends on the viscosity of the fluid and the size of the object. But there is also a buoyant force that depends on the density of the object relative to the fluid. Terminal speed will be greatest for low-viscosity fluids and objects with high densities and small sizes. Thus a skydiver falls more slowly with outspread limbs than when they are in a pike position—head first with hands at their side and legs together.

Note:

Take-Home Experiment: Don’t Lose Your Marbles

By measuring the terminal speed of a slowly moving sphere in a viscous fluid, one can find the viscosity of that fluid (at that temperature). It can be difficult to find small ball bearings around the house, but a small marble will do. Gather two or three fluids (syrup, motor oil, honey, olive oil, etc.) and a thick, tall clear glass or vase. Drop the marble into the center of the fluid and time its fall (after letting it drop a little to reach its terminal speed). Compare your values for the terminal speed and see if they are inversely proportional to the viscosities as listed in [link]. Does it make a difference if the marble is dropped near the side of the glass?

Knowledge of terminal speed is useful for estimating sedimentation rates of small particles. We know from watching mud settle out of dirty water that sedimentation is usually a slow process. Centrifuges are used to speed sedimentation by creating accelerated frames in which gravitational acceleration is replaced by centripetal acceleration, which can be much greater, increasing the terminal speed.


Free body diagram
$\mathbf{F}_v$
$\mathbf{F}_B$
$\mathbf{F}_v$
$\mathbf{F}_B$
L
ρ_{obj}
ρ_{fl}
w
w

There are three forces acting on an object falling through a viscous fluid: its weight w, the viscous drag F_V, and the buoyant force $\mathbf{F}_B$.

Section Summary

- When an object moves in a fluid, there is a different form of the Reynolds number $N\prime_R = \frac{\rho vL}{\eta}\text{(object in fluid)}$, which indicates whether flow is laminar or turbulent.
- For $N\prime_R$ less than about one, flow is laminar.
- For $N\prime_R$ greater than 10^6, flow is entirely turbulent.

Conceptual Questions

Exercise:

Problem:

What direction will a helium balloon move inside a car that is slowing down—toward the front or back? Explain your answer.

Exercise:

Problem:

Will identical raindrops fall more rapidly in $5^{\circ}\ \mathrm{C}$ air or $25^{\circ}\ \mathrm{C}$ air, neglecting any differences in air density? Explain your answer.

Exercise:

Problem:

If you took two marbles of different sizes, what would you expect to observe about the relative magnitudes of their terminal velocities?

Glossary

viscous drag
: a resistance force exerted on a moving object, with a nontrivial dependence on velocity

terminal speed
: the speed at which the viscous drag of an object falling in a viscous fluid is equal to the other forces acting on the object (such as gravity), so that the acceleration of the object is zero

Molecular Transport Phenomena: Diffusion, Osmosis, and Related Processes

- Define diffusion, osmosis, dialysis, and active transport.
- Calculate diffusion rates.

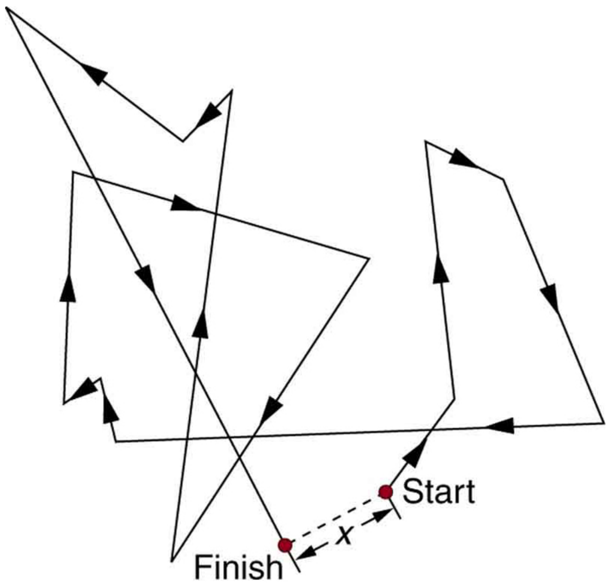
Diffusion

There is something fishy about the ice cube from your freezer—how did it pick up those food odors? How does soaking a sprained ankle in Epsom salt reduce swelling? The answer to these questions are related to atomic and molecular transport phenomena—another mode of fluid motion. Atoms and molecules are in constant motion at any temperature. In fluids they move about randomly even in the absence of macroscopic flow. This motion is called a random walk and is illustrated in [link]. **Diffusion** is the movement of substances due to random thermal molecular motion. Fluids, like fish fumes or odors entering ice cubes, can even diffuse through solids.

Diffusion is a slow process over macroscopic distances. The densities of common materials are great enough that molecules cannot travel very far before having a collision that can scatter them in any direction, including straight backward. It can be shown that the average distance x_{rms} that a molecule travels is proportional to the square root of time:
Equation:

$$x_{\text{rms}} = \sqrt{2\text{Dt}},$$

where x_{rms} stands for the **root-mean-square distance** and is the statistical average for the process. The quantity D is the diffusion constant for the particular molecule in a specific medium. [link] lists representative values of D for various substances, in units of m^2/s.



The random thermal motion of a molecule in a fluid in time t. This type of motion is called a random walk.

Diffusing molecule	**Medium**	***D*** **(m^2/s)**
Hydrogen (H_2)	Air	6.4×10^{-5}
Oxygen (O_2)	Air	1.8×10^{-5}
Oxygen (O_2)	Water	1.0×10^{-9}
Glucose $(C_6H_{12}O_6)$	Water	6.7×10^{-10}
Hemoglobin	Water	6.9×10^{-11}

Diffusing molecule	Medium	*D* (m^2/s)
DNA	Water	1.3×10^{-12}

Diffusion Constants for Various Molecules[footnote]
At 20°C and 1 atm

Note that D gets progressively smaller for more massive molecules. This decrease is because the average molecular speed at a given temperature is inversely proportional to molecular mass. Thus the more massive molecules diffuse more slowly. Another interesting point is that D for oxygen in air is much greater than D for oxygen in water. In water, an oxygen molecule makes many more collisions in its random walk and is slowed considerably. In water, an oxygen molecule moves only about 40 μm in 1 s. (Each molecule actually collides about 10^{10} times per second!). Finally, note that diffusion constants increase with temperature, because average molecular speed increases with temperature. This is because the average kinetic energy of molecules, $\frac{1}{2}mv^2$, is proportional to absolute temperature.

Example:
Calculating Diffusion: How Long Does Glucose Diffusion Take?
Calculate the average time it takes a glucose molecule to move 1.0 cm in water.

Strategy

We can use $x_{\text{rms}} = \sqrt{2Dt}$, the expression for the average distance moved in time t, and solve it for t. All other quantities are known.

Solution

Solving for t and substituting known values yields

Equation:

$$\begin{aligned} t &= \frac{x_{\text{rms}}^2}{2D} = \frac{(0.010\ \text{m})^2}{2(6.7\times10^{-10}\ \text{m}^2/\text{s})} \\ &= 7.5 \times 10^4\ \text{s} = 21\ \text{h}. \end{aligned}$$

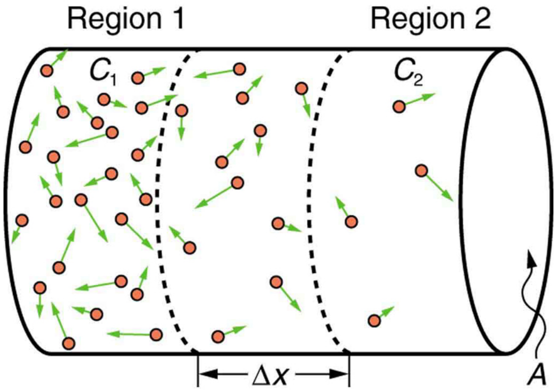
Discussion

This is a remarkably long time for glucose to move a mere centimeter! For this reason, we stir sugar into water rather than waiting for it to diffuse.

Because diffusion is typically very slow, its most important effects occur over small distances. For example, the cornea of the eye gets most of its oxygen by diffusion through the thin tear layer covering it.

The Rate and Direction of Diffusion

If you very carefully place a drop of food coloring in a still glass of water, it will slowly diffuse into the colorless surroundings until its concentration is the same everywhere. This type of diffusion is called free diffusion, because there are no barriers inhibiting it. Let us examine its direction and rate. Molecular motion is random in direction, and so simple chance dictates that more molecules will move out of a region of high concentration than into it. The net rate of diffusion is higher initially than after the process is partially completed. (See [link].)



Diffusion proceeds from a region of higher concentration to a lower one. The net rate of movement is proportional to the difference in concentration.

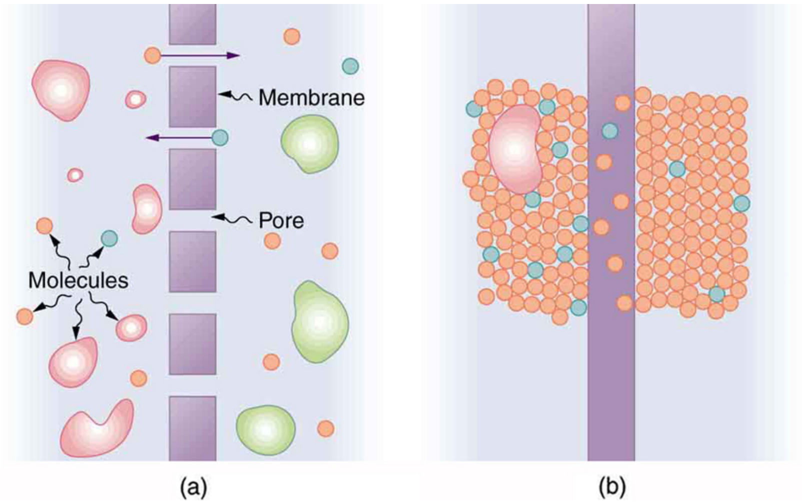
The net rate of diffusion is proportional to the concentration difference. Many more molecules will leave a region of high concentration than will enter it from a region of low concentration. In fact, if the concentrations were the same, there would be *no* net movement. The net rate of diffusion is also proportional to the diffusion constant D, which is determined experimentally. The farther a molecule can diffuse in a given time, the more likely it is to leave the region of high concentration. Many of the factors that affect the rate are hidden in the diffusion constant D. For example, temperature and cohesive and adhesive forces all affect values of D.

Diffusion is the dominant mechanism by which the exchange of nutrients and waste products occur between the blood and tissue, and between air and blood in the lungs. In the evolutionary process, as organisms became larger, they needed quicker methods of transportation than net diffusion, because of the larger distances involved in the transport, leading to the development of circulatory systems. Less sophisticated, single-celled organisms still rely totally on diffusion for the removal of waste products and the uptake of nutrients.

Osmosis and Dialysis—Diffusion across Membranes

Some of the most interesting examples of diffusion occur through barriers that affect the rates of diffusion. For example, when you soak a swollen ankle in Epsom salt, water diffuses through your skin. Many substances regularly move through cell membranes; oxygen moves in, carbon dioxide moves out, nutrients go in, and wastes go out, for example. Because membranes are thin structures (typically 6.5×10^{-9} to 10×10^{-9} m across) diffusion rates through them can be high. Diffusion through membranes is an important method of transport.

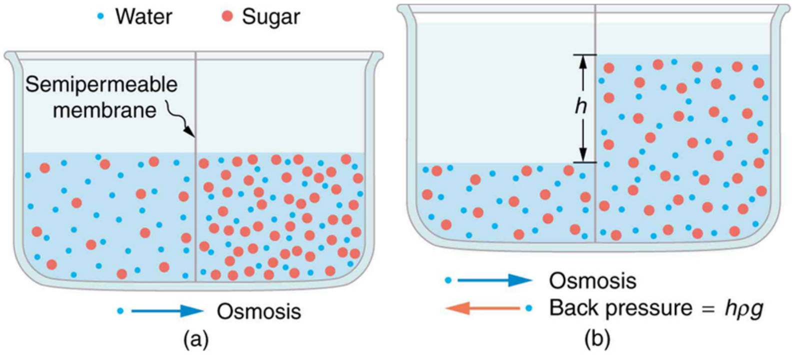
Membranes are generally selectively permeable, or **semipermeable**. (See [link].) One type of semipermeable membrane has small pores that allow only small molecules to pass through. In other types of membranes, the molecules may actually dissolve in the membrane or react with molecules in the membrane while moving across. Membrane function, in fact, is the subject of much current research, involving not only physiology but also chemistry and physics.



(a) A semipermeable membrane with small pores that allow only small molecules to pass through. (b) Certain molecules dissolve in this membrane and diffuse across it.

Osmosis is the transport of water through a semipermeable membrane from a region of high concentration to a region of low concentration. Osmosis is driven by the imbalance in water concentration. For example, water is more concentrated in your body than in Epsom salt. When you soak a swollen ankle in Epsom salt, the water moves out of your body into the lower-concentration region in the salt. Similarly, **dialysis** is the transport of any other molecule through a semipermeable membrane due to its concentration difference. Both osmosis and dialysis are used by the kidneys to cleanse the blood.

Osmosis can create a substantial pressure. Consider what happens if osmosis continues for some time, as illustrated in [link]. Water moves by osmosis from the left into the region on the right, where it is less concentrated, causing the solution on the right to rise. This movement will continue until the pressure ρgh created by the extra height of fluid on the right is large enough to stop further osmosis. This pressure is called a *back pressure*. The back pressure ρgh that stops osmosis is also called the **relative osmotic pressure** if neither solution is pure water, and it is called the **osmotic pressure** if one solution is pure water. Osmotic pressure can be large, depending on the size of the concentration difference. For example, if pure water and sea water are separated by a semipermeable membrane that passes no salt, osmotic pressure will be 25.9 atm. This value means that water will diffuse through the membrane until the salt water surface rises 268 m above the pure-water surface! One example of pressure created by osmosis is turgor in plants (many wilt when too dry). Turgor describes the condition of a plant in which the fluid in a cell exerts a pressure against the cell wall. This pressure gives the plant support. Dialysis can similarly cause substantial pressures.



(a) Two sugar-water solutions of different concentrations, separated by a semipermeable membrane that passes water but not sugar. Osmosis will be to the right, since water is less concentrated there. (b) The fluid level rises until the back pressure ρgh equals the relative osmotic pressure; then, the net transfer of water is zero.

Reverse osmosis and **reverse dialysis** (also called filtration) are processes that occur when back pressure is sufficient to reverse the normal direction of substances through membranes. Back pressure can be created naturally as on the right side of [link]. (A piston can also create this pressure.) Reverse osmosis can be used to desalinate water by simply forcing it through a membrane that will not pass salt. Similarly, reverse dialysis can be used to filter out any substance that a given membrane will not pass.

One further example of the movement of substances through membranes deserves mention. We sometimes find that substances pass in the direction opposite to what we expect. Cypress tree roots, for example, extract pure water from salt water, although osmosis would move it in the opposite direction. This is not reverse osmosis, because there is no back pressure to cause it. What is happening is called **active transport**, a process in which a living membrane expends energy to move substances across it. Many living membranes move water and other substances by active transport. The kidneys, for example, not only use osmosis and dialysis—they also employ significant active transport to move substances into and out of blood. In fact, it is estimated that at least 25% of the body’s energy is expended on active transport of substances at the cellular level. The study of active transport carries us into the realms of microbiology, biophysics, and biochemistry and it is a fascinating application of the laws of nature to living structures.

Section Summary

- Diffusion is the movement of substances due to random thermal molecular motion.
- The average distance x_{rms} a molecule travels by diffusion in a given amount of time is given by
  Equation:

  $$x_{\text{rms}} = \sqrt{2D\text{t}},$$

  where D is the diffusion constant, representative values of which are found in [link].

- Osmosis is the transport of water through a semipermeable membrane from a region of high concentration to a region of low concentration.
- Dialysis is the transport of any other molecule through a semipermeable membrane due to its concentration difference.
- Both processes can be reversed by back pressure.
- Active transport is a process in which a living membrane expends energy to move substances across it.

Conceptual Questions

Exercise:

Problem:

Why would you expect the rate of diffusion to increase with temperature? Can you give an example, such as the fact that you can dissolve sugar more rapidly in hot water?

Exercise:

Problem: How are osmosis and dialysis similar? How do they differ?

Problem Exercises

Exercise:

Problem:

You can smell perfume very shortly after opening the bottle. To show that it is not reaching your nose by diffusion, calculate the average distance a perfume molecule moves in one second in air, given its diffusion constant D to be $1.00 \times 10^{-6}\ \text{m}^2/\text{s}$.

Solution:

$1.41 \times 10^{-3}\ \text{m}$

Exercise:

Problem:

What is the ratio of the average distances that oxygen will diffuse in a given time in air and water? Why is this distance less in water (equivalently, why is D less in water)?

Exercise:

Problem:

Oxygen reaches the veinless cornea of the eye by diffusing through its tear layer, which is 0.500-mm thick. How long does it take the average oxygen molecule to do this?

Solution:

$1.3 \times 10^2\text{ s}$

Exercise:

Problem:

(a) Find the average time required for an oxygen molecule to diffuse through a 0.200-mm-thick tear layer on the cornea. (b) How much time is required to diffuse 0.500 cm^3 of oxygen to the cornea if its surface area is 1.00 cm^2?

Exercise:

Problem:

Suppose hydrogen and oxygen are diffusing through air. A small amount of each is released simultaneously. How much time passes before the hydrogen is 1.00 s ahead of the oxygen? Such differences in arrival times are used as an analytical tool in gas chromatography.

Solution:

0.391 s

Glossary

diffusion
the movement of substances due to random thermal molecular motion

semipermeable
a type of membrane that allows only certain small molecules to pass through

osmosis
the transport of water through a semipermeable membrane from a region of high concentration to one of low concentration

dialysis
the transport of any molecule other than water through a semipermeable membrane from a region of high concentration to one of low concentration

relative osmotic pressure
the back pressure which stops the osmotic process if neither solution is pure water

osmotic pressure
the back pressure which stops the osmotic process if one solution is pure water

reverse osmosis
the process that occurs when back pressure is sufficient to reverse the normal direction of osmosis through membranes

reverse dialysis
the process that occurs when back pressure is sufficient to reverse the normal direction of dialysis through membranes

active transport
the process in which a living membrane expends energy to move substances across

Introduction to the Physics of Hearing class="introduction"

This tree fell some time ago. When it fell, atoms in the air were disturbed. Physicists would call this disturbance sound whether someone was around to hear it or not. (credit: B.A. Bowen Photography)

If a tree falls in the forest and no one is there to hear it, does it make a sound? The answer to this old philosophical question depends on how you define sound. If sound only exists when someone is around to perceive it, then there was no sound. However, if we define sound in terms of physics; that is, a disturbance of the atoms in matter transmitted from its origin outward (in other words, a wave), then there *was* a sound, even if nobody was around to hear it.

Such a wave is the physical phenomenon we call *sound*. Its perception is hearing. Both the physical phenomenon and its perception are interesting and will be considered in this text. We shall explore both sound and hearing; they are related, but are not the same thing. We will also explore the many practical uses of sound waves, such as in medical imaging.

Sound

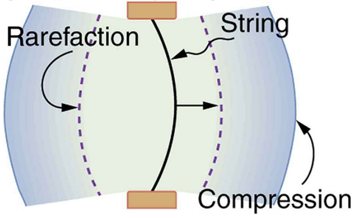
- Define sound and hearing.
- Describe sound as a longitudinal wave.

This glass has been shattered by a high-intensity sound wave of the same frequency as the resonant frequency of the glass. While the sound is not visible, the effects of the sound prove its existence. (credit: ||read||, Flickr)

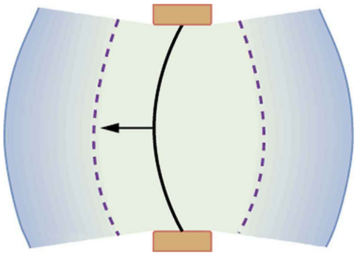
Sound can be used as a familiar illustration of waves. Because hearing is one of our most important senses, it is interesting to see how the physical properties of sound correspond to our perceptions of it. **Hearing** is the perception of sound, just as vision is the perception of visible light. But sound has important applications beyond hearing. Ultrasound, for example, is not heard but can be employed to form medical images and is also used in treatment.

The physical phenomenon of **sound** is defined to be a disturbance of matter that is transmitted from its source outward. Sound is a wave. On the atomic scale, it is a disturbance of atoms that is far more ordered than their thermal motions. In many instances, sound is a periodic wave, and the atoms undergo simple harmonic motion. In this text, we shall explore such periodic sound waves.

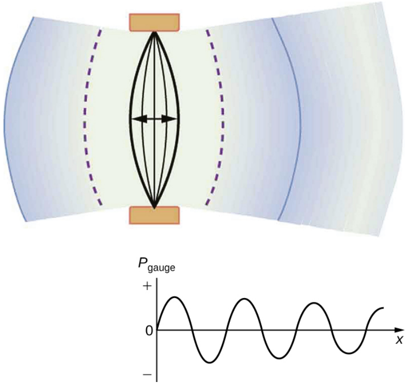
A vibrating string produces a sound wave as illustrated in [link], [link], and [link]. As the string oscillates back and forth, it transfers energy to the air, mostly as thermal energy created by turbulence. But a small part of the string's energy goes into compressing and expanding the surrounding air, creating slightly higher and lower local pressures. These compressions (high pressure regions) and rarefactions (low pressure regions) move out as longitudinal pressure waves having the same frequency as the string—they are the disturbance that is a sound wave. (Sound waves in air and most fluids are longitudinal, because fluids have almost no shear strength. In solids, sound waves can be both transverse and longitudinal.) [link] shows a graph of gauge pressure versus distance from the vibrating string.



A vibrating string moving to the right compresses the air in front of it and expands the air behind it.

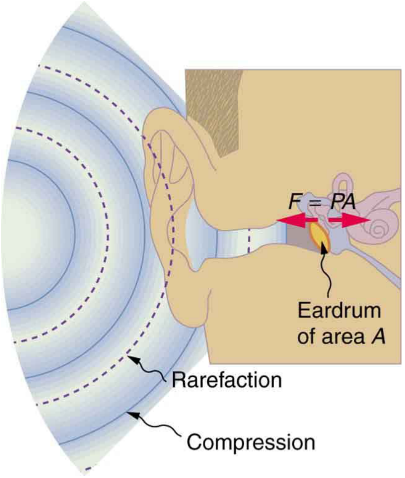


As the string moves to the left, it creates another compression and rarefaction as the ones on the right move away from the string.



After many vibrations, there are a series of compressions and rarefactions moving out from the string as a sound wave. The graph shows gauge pressure versus distance from the source. Pressures vary only slightly from atmospheric for ordinary sounds.

The amplitude of a sound wave decreases with distance from its source, because the energy of the wave is spread over a larger and larger area. But it is also absorbed by objects, such as the eardrum in [link], and converted to thermal energy by the viscosity of air. In addition, during each compression a little heat transfers to the air and during each rarefaction even less heat transfers from the air, so that the heat transfer reduces the organized disturbance into random thermal motions. (These processes can be viewed as a manifestation of the second law of thermodynamics presented in Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency.) Whether the heat transfer from compression to rarefaction is significant depends on how far apart they are—that is, it depends on wavelength. Wavelength, frequency, amplitude, and speed of propagation are important for sound, as they are for all waves.



Sound wave compressions and rarefactions travel up the ear canal and force the eardrum to vibrate. There is a net force on the eardrum, since the sound wave pressures differ from the atmospheric pressure found behind the eardrum. A complicated mechanism converts the vibrations to nerve impulses, which are perceived by the person.

Note:
PhET Explorations: Wave Interference
WMake waves with a dripping faucet, audio speaker, or laser! Add a second source or a pair of slits to create an interference pattern.
https://archive.cnx.org/specials/2fe7ad15-b00e-4402-b068-ff503985a18f/wave-interference/

Section Summary

- Sound is a disturbance of matter that is transmitted from its source outward.
- Sound is one type of wave.

- Hearing is the perception of sound.

Glossary

sound
a disturbance of matter that is transmitted from its source outward

hearing
the perception of sound

Speed of Sound, Frequency, and Wavelength

- Define pitch.
- Describe the relationship between the speed of sound, its frequency, and its wavelength.
- Describe the effects on the speed of sound as it travels through various media.
- Describe the effects of temperature on the speed of sound.

When a firework explodes, the light energy is perceived before the sound energy. Sound travels more slowly than light does. (credit: Dominic Alves, Flickr)

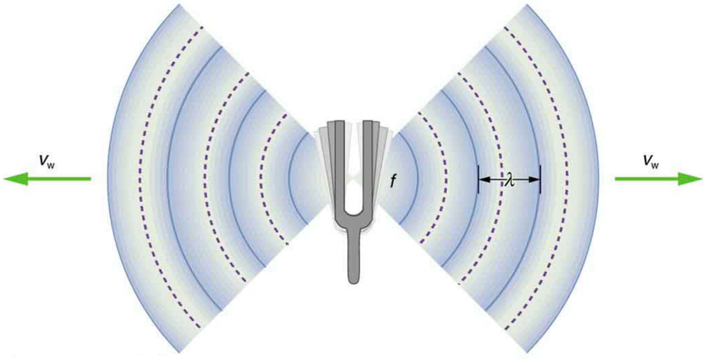
Sound, like all waves, travels at a certain speed and has the properties of frequency and wavelength. You can observe direct evidence of the speed of sound while watching a fireworks display. The flash of an explosion is seen well before its sound is heard, implying both that sound travels at a finite speed and that it is much slower than light. You can also directly sense the frequency of a sound. Perception of frequency is called **pitch**. The wavelength of sound is not directly sensed, but indirect evidence is found in the correlation of the size of musical instruments with their pitch. Small instruments, such as a piccolo, typically make high-pitch sounds, while large instruments, such as a tuba, typically make low-pitch sounds. High pitch means small wavelength, and the size of a musical instrument is directly related to the wavelengths of sound it produces. So a small instrument creates short-wavelength sounds. Similar arguments hold that a large instrument creates long-wavelength sounds.

The relationship of the speed of sound, its frequency, and wavelength is the same as for all waves:

Equation:

$$v_{\text{w}} = f\lambda,$$

where v_{w} is the speed of sound, f is its frequency, and λ is its wavelength. The wavelength of a sound is the distance between adjacent identical parts of a wave—for example, between adjacent compressions as illustrated in [link]. The frequency is the same as that of the source and is the number of waves that pass a point per unit time.



A sound wave emanates from a source vibrating at a frequency f, propagates at v_{w}, and has a wavelength λ.

[link] makes it apparent that the speed of sound varies greatly in different media. The speed of sound in a medium is determined by a combination of the medium’s rigidity (or compressibility in gases) and its density. The

more rigid (or less compressible) the medium, the faster the speed of sound. This observation is analogous to the fact that the frequency of a simple harmonic motion is directly proportional to the stiffness of the oscillating object. The greater the density of a medium, the slower the speed of sound. This observation is analogous to the fact that the frequency of a simple harmonic motion is inversely proportional to the mass of the oscillating object. The speed of sound in air is low, because air is compressible. Because liquids and solids are relatively rigid and very difficult to compress, the speed of sound in such media is generally greater than in gases.

Medium	**v_w(m/s)**
Gases at $0^o C$	
Air	331
Carbon dioxide	259
Oxygen	316
Helium	965
Hydrogen	1290
Liquids at $20^o C$	
Ethanol	1160
Mercury	1450
Water, fresh	1480

Medium	**v_w(m/s)**
Sea water	1540
Human tissue	1540
Solids (longitudinal or bulk)	
Vulcanized rubber	54
Polyethylene	920
Marble	3810
Glass, Pyrex	5640
Lead	1960
Aluminum	5120
Steel	5960

Speed of Sound in Various Media

Earthquakes, essentially sound waves in Earth’s crust, are an interesting example of how the speed of sound depends on the rigidity of the medium. Earthquakes have both longitudinal and transverse components, and these travel at different speeds. The bulk modulus of granite is greater than its shear modulus. For that reason, the speed of longitudinal or pressure waves (P-waves) in earthquakes in granite is significantly higher than the speed of transverse or shear waves (S-waves). Both components of earthquakes travel slower in less rigid material, such as sediments. P-waves have speeds of 4 to 7 km/s, and S-waves correspondingly range in speed from 2 to 5 km/s, both being faster in more rigid material. The P-wave gets progressively farther ahead of the S-wave as they travel through Earth’s crust. The time between the P- and S-waves is routinely used to determine the distance to their source, the epicenter of the earthquake.

The speed of sound is affected by temperature in a given medium. For air at sea level, the speed of sound is given by

Equation:

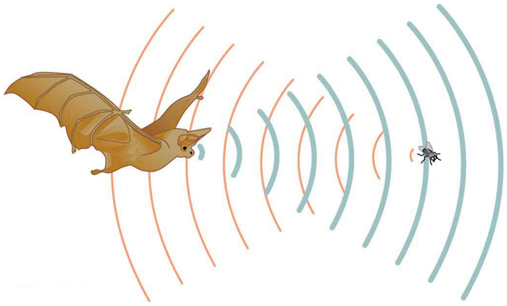
$$v_{\text{w}} = (331\text{ m/s})\sqrt{\frac{T}{273\text{ K}}},$$

where the temperature (denoted as T) is in units of kelvin. The speed of sound in gases is related to the average speed of particles in the gas, v_{rms}, and that

Equation:

$$v_{\text{rms}} = \sqrt{\frac{3\,kT}{m}},$$

where k is the Boltzmann constant ($1.38 \times 10^{-23}\text{ J/K}$) and m is the mass of each (identical) particle in the gas. So, it is reasonable that the speed of sound in air and other gases should depend on the square root of temperature. While not negligible, this is not a strong dependence. At 0ºC, the speed of sound is 331 m/s, whereas at 20.0ºC it is 343 m/s, less than a 4% increase. [link] shows a use of the speed of sound by a bat to sense distances. Echoes are also used in medical imaging.



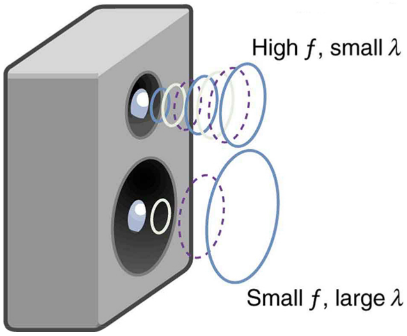
A bat uses sound echoes to find its way about and to catch prey. The time for the echo to return is directly proportional to the distance.

One of the more important properties of sound is that its speed is nearly independent of frequency. This independence is certainly true in open air for sounds in the audible range of 20 to 20,000 Hz. If this independence were not true, you would certainly notice it for music played by a marching band in a football stadium, for example. Suppose that high-frequency sounds traveled faster—then the farther you were from the band, the more the sound from the low-pitch instruments would lag that from the high-pitch ones. But the music from all instruments arrives in cadence independent of distance, and so all frequencies must travel at nearly the same speed. Recall that

Equation:

$$v_{\text{w}} = f\lambda.$$

In a given medium under fixed conditions, v_{w} is constant, so that there is a relationship between f and λ; the higher the frequency, the smaller the wavelength. See [link] and consider the following example.



Because they travel at the same speed in a given medium, low-frequency sounds must have a greater wavelength than high-frequency sounds.

Here, the lower-frequency sounds are emitted by the large speaker, called a woofer, while the higher-frequency sounds are emitted by the small speaker, called a tweeter.

Example:

Calculating Wavelengths: What Are the Wavelengths of Audible Sounds?

Calculate the wavelengths of sounds at the extremes of the audible range, 20 and 20,000 Hz, in 30.0°C air. (Assume that the frequency values are accurate to two significant figures.)

Strategy

To find wavelength from frequency, we can use $v_\text{w} = f\lambda$.

Solution

1. Identify knowns. The value for v_w, is given by
   Equation:

   $$v_\text{w} = (331\ \text{m/s})\sqrt{\frac{T}{273\ \text{K}}}.$$

2. Convert the temperature into kelvin and then enter the temperature into the equation
   Equation:

   $$v_\text{w} = (331\ \text{m/s})\sqrt{\frac{303\ \text{K}}{273\ \text{K}}} = 348.7\ \text{m/s}.$$

3. Solve the relationship between speed and wavelength for λ:
   Equation:

$$\lambda = \frac{v_{\text{w}}}{f}.$$

4. Enter the speed and the minimum frequency to give the maximum wavelength:
   Equation:

$$\lambda_{\text{max}} = \frac{348.7 \text{ m/s}}{20 \text{ Hz}} = 17 \text{ m}.$$

5. Enter the speed and the maximum frequency to give the minimum wavelength:
   Equation:

$$\lambda_{\text{min}} = \frac{348.7 \text{ m/s}}{20{,}000 \text{ Hz}} = 0.017 \text{ m} = 1.7 \text{ cm}.$$

Discussion
Because the product of f multiplied by λ equals a constant, the smaller f is, the larger λ must be, and vice versa.

The speed of sound can change when sound travels from one medium to another. However, the frequency usually remains the same because it is like a driven oscillation and has the frequency of the original source. If v_{w} changes and f remains the same, then the wavelength λ must change. That is, because $v_{\text{w}} = f\lambda$, the higher the speed of a sound, the greater its wavelength for a given frequency.

Note:
Making Connections: Take-Home Investigation—Voice as a Sound Wave

Suspend a sheet of paper so that the top edge of the paper is fixed and the bottom edge is free to move. You could tape the top edge of the paper to the edge of a table. Gently blow near the edge of the bottom of the sheet and note how the sheet moves. Speak softly and then louder such that the sounds hit the edge of the bottom of the paper, and note how the sheet moves. Explain the effects.

Exercise:
Check Your Understanding

Problem:

Imagine you observe two fireworks explode. You hear the explosion of one as soon as you see it. However, you see the other firework for several milliseconds before you hear the explosion. Explain why this is so.

Solution:

Sound and light both travel at definite speeds. The speed of sound is slower than the speed of light. The first firework is probably very close by, so the speed difference is not noticeable. The second firework is farther away, so the light arrives at your eyes noticeably sooner than the sound wave arrives at your ears.

Exercise:
Check Your Understanding

Problem:

You observe two musical instruments that you cannot identify. One plays high-pitch sounds and the other plays low-pitch sounds. How could you determine which is which without hearing either of them play?

Solution:

Compare their sizes. High-pitch instruments are generally smaller than low-pitch instruments because they generate a smaller wavelength.

Section Summary

The relationship of the speed of sound v_{w}, its frequency f, and its wavelength λ is given by

Equation:

$$v_{\text{w}} = f\lambda,$$

which is the same relationship given for all waves.

In air, the speed of sound is related to air temperature T by

Equation:

$$v_{\text{w}} = (331 \text{ m/s})\sqrt{\frac{T}{273 \text{ K}}}.$$

v_{w} is the same for all frequencies and wavelengths.

Conceptual Questions

Exercise:

Problem:

How do sound vibrations of atoms differ from thermal motion?

Exercise:

Problem:

When sound passes from one medium to another where its propagation speed is different, does its frequency or wavelength change? Explain your answer briefly.

Problems & Exercises

Exercise:

Problem:

When poked by a spear, an operatic soprano lets out a 1200-Hz shriek. What is its wavelength if the speed of sound is 345 m/s?

Solution:

0.288 m

Exercise:

Problem:

What frequency sound has a 0.10-m wavelength when the speed of sound is 340 m/s?

Exercise:

Problem:

Calculate the speed of sound on a day when a 1500 Hz frequency has a wavelength of 0.221 m.

Solution:

332 m/s

Exercise:

Problem:

(a) What is the speed of sound in a medium where a 100-kHz frequency produces a 5.96-cm wavelength? (b) Which substance in [link] is this likely to be?

Exercise:

Problem:

Show that the speed of sound in 20.0^{o}C air is 343 m/s, as claimed in the text.

Solution:
Equation:

$$\begin{aligned} v_{\text{w}} &= (331\ \text{m/s})\sqrt{\frac{T}{273\ \text{K}}} = (331\ \text{m/s})\sqrt{\frac{293\ \text{K}}{273\ \text{K}}} \\ &= 343\ \text{m/s} \end{aligned}$$

Exercise:

Problem:

Air temperature in the Sahara Desert can reach 56.0^{o}C (about 134^{o}F). What is the speed of sound in air at that temperature?

Exercise:

Problem:

Dolphins make sounds in air and water. What is the ratio of the wavelength of a sound in air to its wavelength in seawater? Assume air temperature is 20.0^{o}C.

Solution:

0.223

Exercise:

Problem:

A sonar echo returns to a submarine 1.20 s after being emitted. What is the distance to the object creating the echo? (Assume that the submarine is in the ocean, not in fresh water.)

Exercise:

Problem:

(a) If a submarine’s sonar can measure echo times with a precision of 0.0100 s, what is the smallest difference in distances it can detect? (Assume that the submarine is in the ocean, not in fresh water.)

(b) Discuss the limits this time resolution imposes on the ability of the sonar system to detect the size and shape of the object creating the echo.

Solution:

(a) 7.70 m

(b) This means that sonar is good for spotting and locating large objects, but it isn’t able to resolve smaller objects, or detect the detailed shapes of objects. Objects like ships or large pieces of airplanes can be found by sonar, while smaller pieces must be found by other means.

Exercise:

Problem:

A physicist at a fireworks display times the lag between seeing an explosion and hearing its sound, and finds it to be 0.400 s. (a) How far away is the explosion if air temperature is 24.0°C and if you neglect the time taken for light to reach the physicist? (b) Calculate the distance to the explosion taking the speed of light into account. Note that this distance is negligibly greater.

Exercise:

Problem:

Suppose a bat uses sound echoes to locate its insect prey, 3.00 m away. (See [link].) (a) Calculate the echo times for temperatures of 5.00°C and 35.0°C. (b) What percent uncertainty does this cause for the bat in locating the insect? (c) Discuss the significance of this uncertainty and whether it could cause difficulties for the bat. (In practice, the bat continues to use sound as it closes in, eliminating most of any difficulties imposed by this and other effects, such as motion of the prey.)

Solution:

(a) 18.0 ms, 17.1 ms

(b) 5.00%

(c) This uncertainty could definitely cause difficulties for the bat, if it didn’t continue to use sound as it closed in on its prey. A 5% uncertainty could be the difference between catching the prey around the neck or around the chest, which means that it could miss grabbing its prey.

Glossary

pitch	the perception of the frequency of a sound

Sound Intensity and Sound Level

- Define intensity, sound intensity, and sound pressure level.
- Calculate sound intensity levels in decibels (dB).

Noise on crowded roadways like this one in Delhi makes it hard to hear others unless they shout. (credit: Lingaraj G J, Flickr)

In a quiet forest, you can sometimes hear a single leaf fall to the ground. After settling into bed, you may hear your blood pulsing through your ears. But when a passing motorist has his stereo turned up, you cannot even hear what the person next to you in your car is saying. We are all very familiar with the loudness of sounds and aware that they are related to how energetically the source is vibrating. In cartoons depicting a screaming person (or an animal making a loud noise), the cartoonist often shows an open mouth with a vibrating uvula, the hanging tissue at the back of the mouth, to suggest a loud sound coming from the throat [link]. High noise exposure is hazardous to hearing, and it is common for musicians to have hearing losses that are sufficiently severe that they interfere with the musicians’ abilities to perform. The relevant physical quantity is sound intensity, a concept that is valid for all sounds whether or not they are in the audible range.

Intensity is defined to be the power per unit area carried by a wave. Power is the rate at which energy is transferred by the wave. In equation form, **intensity** I is

Equation:

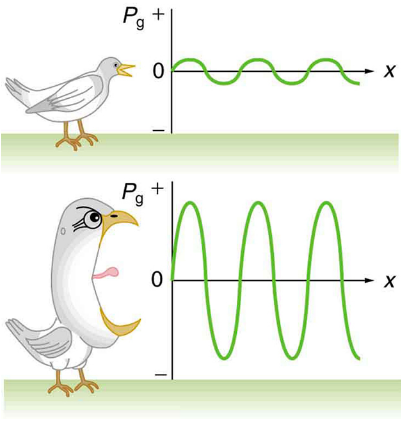
$$I=\frac{P}{A},$$

where P is the power through an area A. The SI unit for I is $\mathrm{W/m^2}$. The intensity of a sound wave is related to its amplitude squared by the following relationship:

Equation:

$$I=\frac{(\Delta p)^2}{2\rho v_{\mathrm{w}}}.$$

Here Δp is the pressure variation or pressure amplitude (half the difference between the maximum and minimum pressure in the sound wave) in units of pascals (Pa) or $\mathrm{N/m^2}$. (We are using a lower case p for pressure to distinguish it from power, denoted by P above.) The energy (as kinetic energy $\frac{mv^2}{2}$) of an oscillating element of air due to a traveling sound wave is proportional to its amplitude squared. In this equation, ρ is the density of the material in which the sound wave travels, in units of $\mathrm{kg/m^3}$, and v_{w} is the speed of sound in the medium, in units of m/s. The pressure variation is proportional to the amplitude of the oscillation, and so I varies as $(\Delta p)^2$ ([link]). This relationship is consistent with the fact that the sound wave is produced by some vibration; the greater its pressure amplitude, the more the air is compressed in the sound it creates.



Graphs of the gauge pressures in two sound waves of different intensities. The more intense sound is produced by a source that has larger-amplitude oscillations and has greater pressure maxima and minima. Because pressures are higher in the greater-intensity sound, it can exert larger forces on the objects it encounters.

Sound intensity levels are quoted in decibels (dB) much more often than sound intensities in watts per meter squared. Decibels are the unit of choice in the scientific literature as well as in the popular media. The reasons for this choice of units are related to how we perceive sounds. How our ears perceive sound can be more accurately described by the logarithm of the intensity rather than directly to the intensity. The **sound intensity level** β in decibels of a sound having an intensity I in watts per meter squared is defined to be

Equation:

$$\beta\;(\text{dB}) = 10\;\log_{10}\left(\frac{I}{I_0}\right),$$

where $I_0 = 10^{-12}\;\text{W/m}^2$ is a reference intensity. In particular, I_0 is the lowest or threshold intensity of sound a person with normal hearing can perceive at a frequency of 1000 Hz. Sound intensity level is not the same as intensity. Because β is defined in terms of a ratio, it is a unitless quantity telling you the *level* of the sound relative to a fixed standard ($10^{-12}\;\text{W/m}^2$, in this case). The units of decibels (dB) are used to indicate this ratio is multiplied by 10 in its definition. The bel, upon which the decibel is based, is named for Alexander Graham Bell, the inventor of the telephone.

Sound intensity level β (dB)	**Intensity I(W/m^2)**	**Example/effect**
0	1×10^{-12}	Threshold of hearing at 1000 Hz
10	1×10^{-11}	Rustle of leaves
20	1×10^{-10}	Whisper at 1 m distance
30	1×10^{-9}	Quiet home

Sound intensity level β (dB)	**Intensity $I(\mathrm{W/m^2})$**	**Example/effect**
40	1×10^{-8}	Average home
50	1×10^{-7}	Average office, soft music
60	1×10^{-6}	Normal conversation
70	1×10^{-5}	Noisy office, busy traffic
80	1×10^{-4}	Loud radio, classroom lecture
90	1×10^{-3}	Inside a heavy truck; damage from prolonged exposure[footnote] Several government agencies and health-related professional associations recommend that 85 dB not be exceeded for 8-hour daily exposures in the absence of hearing protection.
100	1×10^{-2}	Noisy factory, siren at 30 m; damage from 8 h per day exposure
110	1×10^{-1}	Damage from 30 min per day exposure
120	1	Loud rock concert, pneumatic chipper at 2 m; threshold of pain
140	1×10^{2}	Jet airplane at 30 m; severe pain, damage in seconds
160	1×10^{4}	Bursting of eardrums

Sound Intensity Levels and Intensities

Sound Intensity Levels and Intensities

The decibel level of a sound having the threshold intensity of $10^{-12}\ \text{W/m}^2$ is $\beta = 0\ \text{dB}$, because $\log_{10}1 = 0$. That is, the threshold of hearing is 0 decibels. [link] gives levels in decibels and intensities in watts per meter squared for some familiar sounds.

One of the more striking things about the intensities in [link] is that the intensity in watts per meter squared is quite small for most sounds. The ear is sensitive to as little as a trillionth of a watt per meter squared—even more impressive when you realize that the area of the eardrum is only about $1\ \text{cm}^2$, so that only 10^{-16} W falls on it at the threshold of hearing! Air molecules in a sound wave of this intensity vibrate over a distance of less than one molecular diameter, and the gauge pressures involved are less than 10^{-9} atm.

Another impressive feature of the sounds in [link] is their numerical range. Sound intensity varies by a factor of 10^{12} from threshold to a sound that causes damage in seconds. You are unaware of this tremendous range in sound intensity because how your ears respond can be described approximately as the logarithm of intensity. Thus, sound intensity levels in decibels fit your experience better than intensities in watts per meter squared. The decibel scale is also easier to relate to because most people are more accustomed to dealing with numbers such as 0, 53, or 120 than numbers such as 1.00×10^{-11}.

One more observation readily verified by examining [link] or using $I = \frac{(\Delta p)^2}{2\rho v_\text{w}}$ is that each factor of 10 in intensity corresponds to 10 dB. For example, a 90 dB sound compared with a 60 dB sound is 30 dB greater, or three factors of 10 (that is, 10^3 times) as intense. Another example is that if one sound is 10^7 as intense as another, it is 70 dB higher. See [link].

I_2/I_1	$\beta_2 - \beta_1$
2.0	3.0 dB
5.0	7.0 dB
10.0	10.0 dB

Ratios of Intensities and Corresponding Differences in Sound Intensity Levels

Example:

Calculating Sound Intensity Levels: Sound Waves

Calculate the sound intensity level in decibels for a sound wave traveling in air at 0ºC and having a pressure amplitude of 0.656 Pa.

Strategy

We are given Δp, so we can calculate I using the equation $I = (\Delta p)^2/(2pv_{\text{w}})^2$. Using I, we can calculate β straight from its definition in $\beta\ (\text{dB}) = 10\ \log_{10}(I/I_0)$.

Solution

(1) Identify knowns:

Sound travels at 331 m/s in air at 0ºC.

Air has a density of $1.29\ \text{kg/m}^3$ at atmospheric pressure and 0ºC.

(2) Enter these values and the pressure amplitude into $I = (\Delta p)^2/\ (2\rho v_{\text{w}})$:

Equation:

$$I = \frac{(\Delta p)^2}{2\rho v_{\text{w}}} = \frac{(0.656\ \text{Pa})^2}{2\left(1.29\ \text{kg/m}^3\right)(331\ \text{m/s})} = 5.04 \times 10^{-4}\ \text{W/m}^2.$$

(3) Enter the value for I and the known value for I_0 into $\beta\ (\text{dB}) = 10\ \log_{10}(I/I_0)$. Calculate to find the sound intensity level in decibels:

Equation:

$$10\ \log_{10}\left(5.04 \times 10^{8}\right) = 10\ \left(8.70\right)\ \text{dB} = 87\ \text{dB}.$$

Discussion

This 87 dB sound has an intensity five times as great as an 80 dB sound. So a factor of five in intensity corresponds to a difference of 7 dB in sound intensity level. This value is true for any intensities differing by a factor of five.

Example:

Change Intensity Levels of a Sound: What Happens to the Decibel Level?

Show that if one sound is twice as intense as another, it has a sound level about 3 dB higher.

Strategy

You are given that the ratio of two intensities is 2 to 1, and are then asked to find the difference in their sound levels in decibels. You can solve this problem using of the properties of logarithms.

Solution

(1) Identify knowns:

The ratio of the two intensities is 2 to 1, or:

Equation:

$$\frac{I_2}{I_1} = 2.00.$$

We wish to show that the difference in sound levels is about 3 dB. That is, we want to show:

Equation:

$$\beta_2 - \beta_1 = 3\ \text{dB}.$$

Note that:

Equation:

$$\log_{10} b - \log_{10} a = \log_{10}\left(\frac{b}{a}\right).$$

(2) Use the definition of β to get:

Equation:

$$\beta_2 - \beta_1 = 10\ \log_{10}\left(\frac{I_2}{I_1}\right) = 10\ \log_{10} 2.00 = 10\ (0.301)\ \text{dB}.$$

Thus,

Equation:

$$\beta_2 - \beta_1 = 3.01\ \text{dB}.$$

Discussion

This means that the two sound intensity levels differ by 3.01 dB, or about 3 dB, as advertised. Note that because only the ratio I_2/I_1 is given (and not the actual intensities), this result is true for any intensities that differ by a factor of two. For example, a 56.0 dB sound is twice as intense as a 53.0 dB sound, a 97.0 dB sound is half as intense as a 100 dB sound, and so on.

It should be noted at this point that there is another decibel scale in use, called the **sound pressure level**, based on the ratio of the pressure amplitude to a reference pressure. This scale is used particularly in applications where sound travels in water. It is beyond the scope of most introductory texts to treat this scale because it is not commonly used for sounds in air, but it is important to note that very different decibel levels may be encountered when sound pressure levels are quoted. For example, ocean noise pollution produced by ships may be as great as 200 dB expressed in the sound pressure level, where the more familiar sound intensity level we use here would be something under 140 dB for the same sound.

Note:

Take-Home Investigation: Feeling Sound

Find a CD player and a CD that has rock music. Place the player on a light table, insert the CD into the player, and start playing the CD. Place your hand gently on the table next to the speakers. Increase the volume and note the level when the table just begins to vibrate as the rock music plays. Increase the reading on the volume control until it doubles. What has happened to the vibrations?

Exercise:
Check Your Understanding

Problem:

Describe how amplitude is related to the loudness of a sound.

Solution:

Amplitude is directly proportional to the experience of loudness. As amplitude increases, loudness increases.

Exercise:
Check Your Understanding

Problem:

Identify common sounds at the levels of 10 dB, 50 dB, and 100 dB.

Solution:

10 dB: Running fingers through your hair.

50 dB: Inside a quiet home with no television or radio.

100 dB: Take-off of a jet plane.

Section Summary

- Intensity is the same for a sound wave as was defined for all waves; it is
  Equation:

  $$I = \frac{P}{A},$$

  where P is the power crossing area A. The SI unit for I is watts per meter squared. The intensity of a sound wave is also related to the pressure amplitude Δp
  Equation:

  $$I = \frac{(\Delta p)^2}{2\rho v_{\text{w}}},$$

  where ρ is the density of the medium in which the sound wave travels and v_{w} is the speed of sound in the medium.

- Sound intensity level in units of decibels (dB) is
  Equation:

  $$\beta\,(\text{dB}) = 10\,\log_{10}\left(\frac{I}{I_0}\right),$$

  where $I_0 = 10^{-12}\ \text{W/m}^2$ is the threshold intensity of hearing.

Conceptual Questions

Exercise:

Problem:

Six members of a synchronized swim team wear earplugs to protect themselves against water pressure at depths, but they can still hear the music and perform the combinations in the water perfectly. One day, they were asked to leave the pool so the dive team could practice a few dives, and they tried to practice on a mat, but seemed to have a lot more difficulty. Why might this be?

Exercise:

Problem:

A community is concerned about a plan to bring train service to their downtown from the town’s outskirts. The current sound intensity level, even though the rail yard is blocks away, is 70 dB downtown. The mayor assures the public that there will be a difference of only 30 dB in sound in the downtown area. Should the townspeople be concerned? Why?

Problems & Exercises

Exercise:

Problem:

What is the intensity in watts per meter squared of 85.0-dB sound?

Solution:
Equation:

$$3.16 \times 10^{-4}\ \text{W/m}^2$$

Exercise:

Problem:

The warning tag on a lawn mower states that it produces noise at a level of 91.0 dB. What is this in watts per meter squared?

Exercise:

Problem:

A sound wave traveling in 20°C air has a pressure amplitude of 0.5 Pa. What is the intensity of the wave?

Solution:

Equation:

$$3.04 \times 10^{-4}\ \text{W/m}^2$$

Exercise:

Problem:

What intensity level does the sound in the preceding problem correspond to?

Exercise:

Problem:

What sound intensity level in dB is produced by earphones that create an intensity of $4.00 \times 10^{-2}\ \text{W/m}^2$?

Solution:

106 dB

Exercise:

Problem:

Show that an intensity of $10^{-12}\ \text{W/m}^2$ is the same as $10^{-16}\ \text{W/cm}^2$.

Exercise:

Problem:

(a) What is the decibel level of a sound that is twice as intense as a 90.0-dB sound? (b) What is the decibel level of a sound that is one-fifth as intense as a 90.0-dB sound?

Solution:

(a) 93 dB

(b) 83 dB

Exercise:

Problem:

(a) What is the intensity of a sound that has a level 7.00 dB lower than a $4.00 \times 10^{-9}\ \text{W/m}^2$ sound? (b) What is the intensity of a sound that is 3.00 dB higher than a $4.00 \times 10^{-9}\ \text{W/m}^2$ sound?

Exercise:

Problem:

(a) How much more intense is a sound that has a level 17.0 dB higher than another? (b) If one sound has a level 23.0 dB less than another, what is the ratio of their intensities?

Solution:

(a) 50.1

(b) 5.01×10^{-3} or $\frac{1}{200}$

Exercise:

Problem:

People with good hearing can perceive sounds as low in level as –8.00 dB at a frequency of 3000 Hz. What is the intensity of this sound in watts per meter squared?

Exercise:

Problem:

If a large housefly 3.0 m away from you makes a noise of 40.0 dB, what is the noise level of 1000 flies at that distance, assuming interference has a negligible effect?

Solution:

70.0 dB

Exercise:

Problem:

Ten cars in a circle at a boom box competition produce a 120-dB sound intensity level at the center of the circle. What is the average sound intensity level produced there by each stereo, assuming interference effects can be neglected?

Exercise:

Problem:

The amplitude of a sound wave is measured in terms of its maximum gauge pressure. By what factor does the amplitude of a sound wave increase if the sound intensity level goes up by 40.0 dB?

Solution:

100

Exercise:

Problem:

If a sound intensity level of 0 dB at 1000 Hz corresponds to a maximum gauge pressure (sound amplitude) of $10^{-9}\ \text{atm}$, what is the maximum gauge pressure in a 60-dB sound? What is the maximum gauge pressure in a 120-dB sound?

Exercise:

Problem:

An 8-hour exposure to a sound intensity level of 90.0 dB may cause hearing damage. What energy in joules falls on a 0.800-cm-diameter eardrum so exposed?

Solution:
Equation:

$$1.45 \times 10^{-3}\ \text{J}$$

Exercise:

Problem:

(a) Ear trumpets were never very common, but they did aid people with hearing losses by gathering sound over a large area and concentrating it on the smaller area of the eardrum. What decibel increase does an ear trumpet produce if its sound gathering area is $900\ \text{cm}^2$ and the area of the eardrum is $0.500\ \text{cm}^2$, but the trumpet only has an efficiency of 5.00% in transmitting the sound to the eardrum? (b) Comment on the usefulness of the decibel increase found in part (a).

Exercise:

Problem:

Sound is more effectively transmitted into a stethoscope by direct contact than through the air, and it is further intensified by being concentrated on the smaller area of the eardrum. It is reasonable to assume that sound is transmitted into a stethoscope 100 times as effectively compared with transmission though the air. What, then, is the gain in decibels produced by a stethoscope that has a sound gathering area of 15.0 cm^2, and concentrates the sound onto two eardrums with a total area of 0.900 cm^2 with an efficiency of 40.0%?

Solution:

28.2 dB

Exercise:

Problem:

Loudspeakers can produce intense sounds with surprisingly small energy input in spite of their low efficiencies. Calculate the power input needed to produce a 90.0-dB sound intensity level for a 12.0-cm-diameter speaker that has an efficiency of 1.00%. (This value is the sound intensity level right at the speaker.)

Glossary

intensity	the power per unit area carried by a wave
sound intensity level	a unitless quantity telling you the level of the sound relative to a fixed standard
sound pressure level	the ratio of the pressure amplitude to a reference pressure

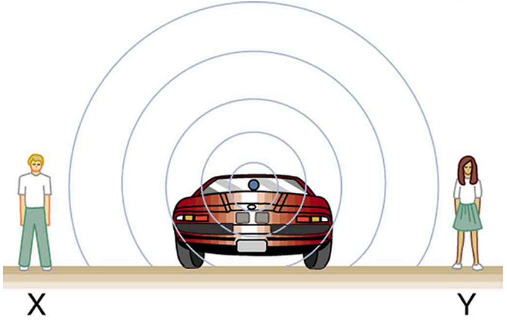
Doppler Effect and Sonic Booms

- Define Doppler effect, Doppler shift, and sonic boom.
- Calculate the frequency of a sound heard by someone observing Doppler shift.
- Describe the sounds produced by objects moving faster than the speed of sound.

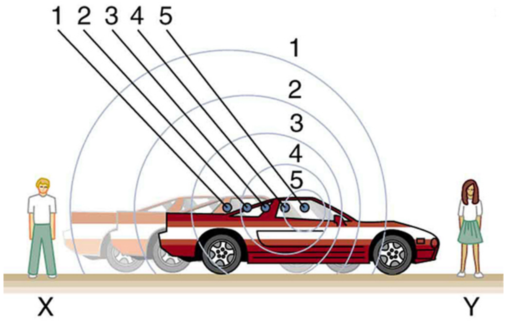
The characteristic sound of a motorcycle buzzing by is an example of the **Doppler effect**. The high-pitch scream shifts dramatically to a lower-pitch roar as the motorcycle passes by a stationary observer. The closer the motorcycle brushes by, the more abrupt the shift. The faster the motorcycle moves, the greater the shift. We also hear this characteristic shift in frequency for passing race cars, airplanes, and trains. It is so familiar that it is used to imply motion and children often mimic it in play.

The Doppler effect is an alteration in the observed frequency of a sound due to motion of either the source or the observer. Although less familiar, this effect is easily noticed for a stationary source and moving observer. For example, if you ride a train past a stationary warning bell, you will hear the bell's frequency shift from high to low as you pass by. The actual change in frequency due to relative motion of source and observer is called a **Doppler shift**. The Doppler effect and Doppler shift are named for the Austrian physicist and mathematician Christian Johann Doppler (1803–1853), who did experiments with both moving sources and moving observers. Doppler, for example, had musicians play on a moving open train car and also play standing next to the train tracks as a train passed by. Their music was observed both on and off the train, and changes in frequency were measured.

What causes the Doppler shift? [link], [link], and [link] compare sound waves emitted by stationary and moving sources in a stationary air mass. Each disturbance spreads out spherically from the point where the sound was emitted. If the source is stationary, then all of the spheres representing the air compressions in the sound wave centered on the same point, and the stationary observers on either side see the same wavelength and frequency as emitted by the source, as in [link]. If the source is moving, as in [link], then the situation is different. Each compression of the air moves out in a sphere from the point where it was emitted, but the point of emission moves. This moving emission point causes the air compressions to be closer together on one side and farther apart on the other. Thus, the wavelength is shorter in the direction the source is moving (on the right in [link]), and longer in the opposite direction (on the left in [link]). Finally, if the observers move, as in [link], the frequency at which they receive the compressions changes. The observer moving toward the source receives them at a higher frequency, and the person moving away from the source receives them at a lower frequency.

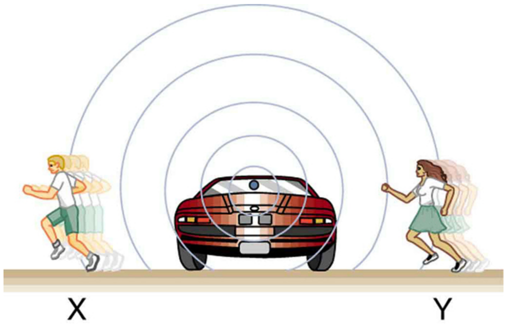


Sounds emitted by a source spread out in spherical waves. Because the source, observers, and air are stationary, the wavelength and frequency are the same in all directions and to all observers.



Sounds emitted by a

source moving to the right spread out from the points at which they were emitted. The wavelength is reduced and, consequently, the frequency is increased in the direction of motion, so that the observer on the right hears a higher-pitch sound. The opposite is true for the observer on the left, where the wavelength is increased and the frequency is reduced.



The same effect is produced when the observers move relative to the source. Motion toward the source increases frequency as the observer on the right passes through more wave crests than she would if stationary. Motion away from the

source decreases
frequency as the observer
on the left passes through
fewer wave crests than he
would if stationary.

We know that wavelength and frequency are related by $v_{\mathrm{w}} = f\lambda$, where v_{w} is the fixed speed of sound. The sound moves in a medium and has the same speed v_{w} in that medium whether the source is moving or not. Thus f multiplied by λ is a constant. Because the observer on the right in [link] receives a shorter wavelength, the frequency she receives must be higher. Similarly, the observer on the left receives a longer wavelength, and hence he hears a lower frequency. The same thing happens in [link]. A higher frequency is received by the observer moving toward the source, and a lower frequency is received by an observer moving away from the source. In general, then, relative motion of source and observer toward one another increases the received frequency. Relative motion apart decreases frequency. The greater the relative speed is, the greater the effect.

Note:
The Doppler Effect
The Doppler effect occurs not only for sound but for any wave when there is relative motion between the observer and the source. There are Doppler shifts in the frequency of sound, light, and water waves, for example. Doppler shifts can be used to determine velocity, such as when ultrasound is reflected from blood in a medical diagnostic. The recession of galaxies is determined by the shift in the frequencies of light received from them and has implied much about the origins of the universe. Modern physics has been profoundly affected by observations of Doppler shifts.

For a stationary observer and a moving source, the frequency f_{obs} received by the observer can be shown to be

Equation:

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} \pm v_{\text{s}}}\right),$$

where f_{s} is the frequency of the source, v_{s} is the speed of the source along a line joining the source and observer, and v_{w} is the speed of sound. The minus sign is used for motion toward the observer and the plus sign for motion away from the observer, producing the appropriate shifts up and down in frequency. Note that the greater the speed of the source, the greater the effect. Similarly, for a stationary source and moving observer, the frequency received by the observer f_{obs} is given by

Equation:

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}} \pm v_{\text{obs}}}{v_{\text{w}}}\right),$$

where v_{obs} is the speed of the observer along a line joining the source and observer. Here the plus sign is for motion toward the source, and the minus is for motion away from the source.

Example:

Calculate Doppler Shift: A Train Horn

Suppose a train that has a 150-Hz horn is moving at 35.0 m/s in still air on a day when the speed of sound is 340 m/s.

(a) What frequencies are observed by a stationary person at the side of the tracks as the train approaches and after it passes?

(b) What frequency is observed by the train's engineer traveling on the train?

Strategy

To find the observed frequency in (a), $f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} \pm v_{\text{s}}}\right)$, must be used because the source is moving. The minus sign is used for the approaching train, and the plus sign for the receding train. In (b), there are two Doppler shifts—one for a moving source and the other for a moving observer.

Solution for (a)

(1) Enter known values into $f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} - v_{\text{s}}}\right)$.

Equation:

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} - v_{\text{s}}}\right) = (150\text{ Hz})\left(\frac{340\text{ m/s}}{340\text{ m/s} – 35.0\text{ m/s}}\right)$$

(2) Calculate the frequency observed by a stationary person as the train approaches.

Equation:

$$f_{\text{obs}} = (150\text{ Hz})(1.11) = 167\text{ Hz}$$

(3) Use the same equation with the plus sign to find the frequency heard by a stationary person as the train recedes.

Equation:

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} + v_{\text{s}}}\right) = (150\text{ Hz})\left(\frac{340\text{ m/s}}{340\text{ m/s} + 35.0\text{ m/s}}\right)$$

(4) Calculate the second frequency.

Equation:

$$f_{\text{obs}} = (150\text{ Hz})(0.907) = 136\text{ Hz}$$

Discussion on (a)

The numbers calculated are valid when the train is far enough away that the motion is nearly along the line joining train and observer. In both cases, the shift is significant and easily noticed. Note that the shift is 17.0 Hz for motion toward and 14.0 Hz for motion away. The shifts are not symmetric.

Solution for (b)

(1) Identify knowns:

- It seems reasonable that the engineer would receive the same frequency as emitted by the horn, because the relative velocity

between them is zero.
- Relative to the medium (air), the speeds are $v_\text{s} = v_\text{obs} = 35.0\text{ m/s}$.
- The first Doppler shift is for the moving observer; the second is for the moving source.

(2) Use the following equation:

Equation:

$$f_\text{obs} = \left[f_\text{s} \left(\frac{v_\text{w} \pm v_\text{obs}}{v_\text{w}} \right) \right] \left(\frac{v_\text{w}}{v_\text{w} \pm v_\text{s}} \right).$$

The quantity in the square brackets is the Doppler-shifted frequency due to a moving observer. The factor on the right is the effect of the moving source.

(3) Because the train engineer is moving in the direction toward the horn, we must use the plus sign for v_obs; however, because the horn is also moving in the direction away from the engineer, we also use the plus sign for v_s. But the train is carrying both the engineer and the horn at the same velocity, so $v_\text{s} = v_\text{obs}$. As a result, everything but f_s cancels, yielding

Equation:

$$f_\text{obs} = f_\text{s}.$$

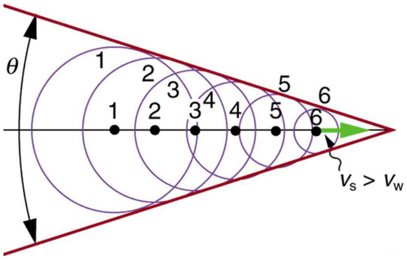
Discussion for (b)

We may expect that there is no change in frequency when source and observer move together because it fits your experience. For example, there is no Doppler shift in the frequency of conversations between driver and passenger on a motorcycle. People talking when a wind moves the air between them also observe no Doppler shift in their conversation. The crucial point is that source and observer are not moving relative to each other.

Sonic Booms to Bow Wakes

What happens to the sound produced by a moving source, such as a jet airplane, that approaches or even exceeds the speed of sound? The answer to this question applies not only to sound but to all other waves as well.

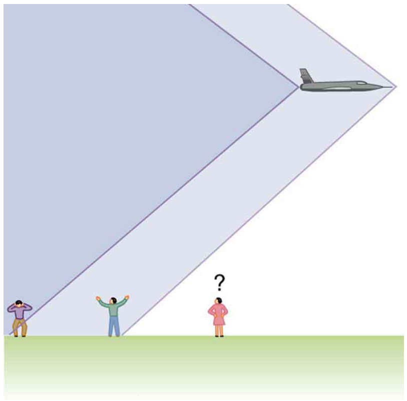
Suppose a jet airplane is coming nearly straight at you, emitting a sound of frequency f_s. The greater the plane's speed v_s, the greater the Doppler shift and the greater the value observed for f_obs. Now, as v_s approaches the speed of sound, f_obs approaches infinity, because the denominator in $f_\text{obs} = f_\text{s}\left(\frac{v_\text{w}}{v_\text{w} \pm v_\text{s}}\right)$ approaches zero. At the speed of sound, this result means that in front of the source, each successive wave is superimposed on the previous one because the source moves forward at the speed of sound. The observer gets them all at the same instant, and so the frequency is infinite. (Before airplanes exceeded the speed of sound, some people argued it would be impossible because such constructive superposition would produce pressures great enough to destroy the airplane.) If the source exceeds the speed of sound, no sound is received by the observer until the source has passed, so that the sounds from the approaching source are mixed with those from it when receding. This mixing appears messy, but something interesting happens—a sonic boom is created. (See [link].)



Sound waves from a source that moves faster than the speed of sound spread spherically from the point where they are emitted, but the source moves ahead of each.

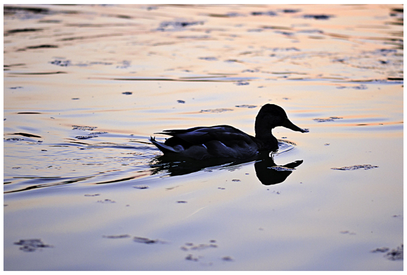
Constructive interference along the lines shown (actually a cone in three dimensions) creates a shock wave called a sonic boom. The faster the speed of the source, the smaller the angle θ.

There is constructive interference along the lines shown (a cone in three dimensions) from similar sound waves arriving there simultaneously. This superposition forms a disturbance called a **sonic boom**, a constructive interference of sound created by an object moving faster than sound. Inside the cone, the interference is mostly destructive, and so the sound intensity there is much less than on the shock wave. An aircraft creates two sonic booms, one from its nose and one from its tail. (See [link].) During television coverage of space shuttle landings, two distinct booms could often be heard. These were separated by exactly the time it would take the shuttle to pass by a point. Observers on the ground often do not see the aircraft creating the sonic boom, because it has passed by before the shock wave reaches them, as seen in [link]. If the aircraft flies close by at low altitude, pressures in the sonic boom can be destructive and break windows as well as rattle nerves. Because of how destructive sonic booms can be, supersonic flights are banned over populated areas of the United States.

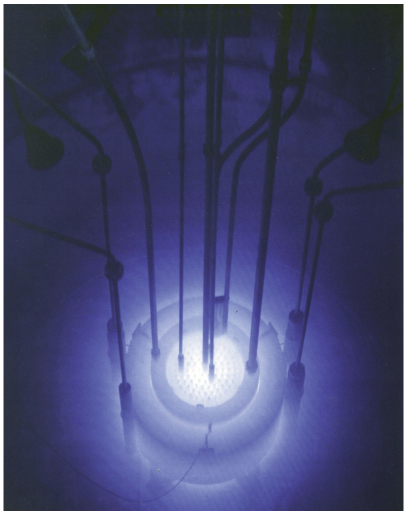


Two sonic booms, created by the nose and tail of an aircraft, are observed on the ground after the plane has passed by.

Sonic booms are one example of a broader phenomenon called bow wakes. A **bow wake**, such as the one in [link], is created when the wave source moves faster than the wave propagation speed. Water waves spread out in circles from the point where created, and the bow wake is the familiar V-shaped wake trailing the source. A more exotic bow wake is created when a subatomic particle travels through a medium faster than the speed of light travels in that medium. (In a vacuum, the maximum speed of light will be $c=3.00\times 10^{8}\ \text{m/s}$; in the medium of water, the speed of light is closer to $0.75c$. If the particle creates light in its passage, that light spreads on a cone with an angle indicative of the speed of the particle, as illustrated in [link]. Such a bow wake is called Cerenkov radiation and is commonly observed in particle physics.



Bow wake created by a duck. Constructive interference produces the rather structured wake, while there is relatively little wave action inside the wake, where interference is mostly destructive. (credit: Horia Varlan, Flickr)



The blue glow in this research reactor pool is Cerenkov radiation caused by subatomic particles traveling faster than the speed of light in water. (credit: U.S. Nuclear Regulatory Commission)

Doppler shifts and sonic booms are interesting sound phenomena that occur in all types of waves. They can be of considerable use. For example, the Doppler shift in ultrasound can be used to measure blood velocity, while police use the Doppler shift in radar (a microwave) to measure car velocities. In meteorology, the Doppler shift is used to track the motion of storm clouds; such “Doppler Radar” can give velocity and direction and rain or snow potential of imposing weather fronts. In astronomy, we can examine the light emitted from distant galaxies and determine their speed relative to ours. As galaxies move away from us, their light is shifted to a lower frequency, and so to a longer wavelength—the so-called red shift. Such information from galaxies far, far away has allowed us to estimate the age of the universe (from the Big Bang) as about 14 billion years.

Exercise:
Check Your Understanding

Problem:

Why did scientist Christian Doppler observe musicians both on a moving train and also from a stationary point not on the train?

Solution:

Doppler needed to compare the perception of sound when the observer is stationary and the sound source moves, as well as when the sound source and the observer are both in motion.

Exercise:
Check Your Understanding

Problem:

Describe a situation in your life when you might rely on the Doppler shift to help you either while driving a car or walking near traffic.

Solution:

If I am driving and I hear Doppler shift in an ambulance siren, I would be able to tell when it was getting closer and also if it has passed by. This would help me to know whether I needed to pull over and let the ambulance through.

Section Summary

- The Doppler effect is an alteration in the observed frequency of a sound due to motion of either the source or the observer.
- The actual change in frequency is called the Doppler shift.
- A sonic boom is constructive interference of sound created by an object moving faster than sound.
- A sonic boom is a type of bow wake created when any wave source moves faster than the wave propagation speed.
- For a stationary observer and a moving source, the observed frequency f_{obs} is:
  Equation:
  $$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} \pm v_{\text{s}}}\right),$$
  where f_{s} is the frequency of the source, v_{s} is the speed of the source, and v_{w} is the speed of sound. The minus sign is used for motion toward the observer and the plus sign for motion away.
- For a stationary source and moving observer, the observed frequency is:

Equation:

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}} \pm v_{\text{obs}}}{v_{\text{w}}}\right),$$

where v_{obs} is the speed of the observer.

Conceptual Questions

Exercise:

Problem: Is the Doppler shift real or just a sensory illusion?

Exercise:

Problem:

Due to efficiency considerations related to its bow wake, the supersonic transport aircraft must maintain a cruising speed that is a constant ratio to the speed of sound (a constant Mach number). If the aircraft flies from warm air into colder air, should it increase or decrease its speed? Explain your answer.

Exercise:

Problem:

When you hear a sonic boom, you often cannot see the plane that made it. Why is that?

Problems & Exercises

Exercise:

Problem:

(a) What frequency is received by a person watching an oncoming ambulance moving at 110 km/h and emitting a steady 800-Hz sound from its siren? The speed of sound on this day is 345 m/s. (b) What frequency does she receive after the ambulance has passed?

Solution:

(a) 878 Hz

(b) 735 Hz

Exercise:

Problem:

(a) At an air show a jet flies directly toward the stands at a speed of 1200 km/h, emitting a frequency of 3500 Hz, on a day when the speed of sound is 342 m/s. What frequency is received by the observers? (b) What frequency do they receive as the plane flies directly away from them?

Exercise:

Problem:

What frequency is received by a mouse just before being dispatched by a hawk flying at it at 25.0 m/s and emitting a screech of frequency 3500 Hz? Take the speed of sound to be 331 m/s.

Solution:
Equation:

$$3.79 \times 10^3 \text{ Hz}$$

Exercise:

Problem:

A spectator at a parade receives an 888-Hz tone from an oncoming trumpeter who is playing an 880-Hz note. At what speed is the musician approaching if the speed of sound is 338 m/s?

Exercise:

Problem:

A commuter train blows its 200-Hz horn as it approaches a crossing. The speed of sound is 335 m/s. (a) An observer waiting at the crossing receives a frequency of 208 Hz. What is the speed of the train? (b) What frequency does the observer receive as the train moves away?

Solution:

(a) 12.9 m/s

(b) 193 Hz

Exercise:

Problem:

Can you perceive the shift in frequency produced when you pull a tuning fork toward you at 10.0 m/s on a day when the speed of sound is 344 m/s? To answer this question, calculate the factor by which the frequency shifts and see if it is greater than 0.300%.

Exercise:

Problem:

Two eagles fly directly toward one another, the first at 15.0 m/s and the second at 20.0 m/s. Both screech, the first one emitting a frequency of 3200 Hz and the second one emitting a frequency of 3800 Hz. What frequencies do they receive if the speed of sound is 330 m/s?

Solution:

First eagle hears $4.23 \times 10^3\ \mathrm{Hz}$

Second eagle hears $3.56 \times 10^3\ \mathrm{Hz}$

Exercise:

Problem:

What is the minimum speed at which a source must travel toward you for you to be able to hear that its frequency is Doppler shifted? That is, what speed produces a shift of 0.300% on a day when the speed of sound is 331 m/s?

Glossary

Doppler effect
: an alteration in the observed frequency of a sound due to motion of either the source or the observer

Doppler shift
: the actual change in frequency due to relative motion of source and observer

sonic boom
: a constructive interference of sound created by an object moving faster than sound

bow wake
: V-shaped disturbance created when the wave source moves faster than the wave propagation speed

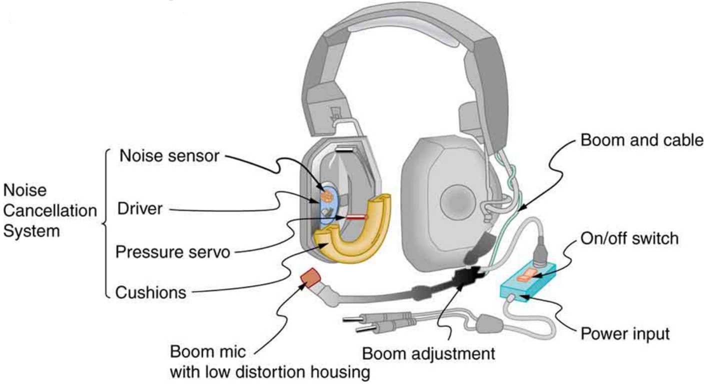
Sound Interference and Resonance: Standing Waves in Air Columns

- Define antinode, node, fundamental, overtones, and harmonics.
- Identify instances of sound interference in everyday situations.
- Describe how sound interference occurring inside open and closed tubes changes the characteristics of the sound, and how this applies to sounds produced by musical instruments.
- Calculate the length of a tube using sound wave measurements.

Some types of headphones use the phenomena of constructive and destructive interference to cancel out outside noises. (credit: JVC America, Flickr)

Interference is the hallmark of waves, all of which exhibit constructive and destructive interference exactly analogous to that seen for water waves. In fact, one way to prove something “is a wave” is to observe interference effects. So, sound being a wave, we expect it to exhibit interference; we have already mentioned a few such effects, such as the beats from two similar notes played simultaneously.

[link] shows a clever use of sound interference to cancel noise. Larger-scale applications of active noise reduction by destructive interference are contemplated for entire passenger compartments in commercial aircraft. To obtain destructive interference, a fast electronic analysis is performed, and a second sound is introduced with its maxima and minima exactly reversed from the incoming noise. Sound waves in fluids are pressure waves and consistent with Pascal’s principle; pressures from two different sources add and subtract like simple numbers; that is, positive and negative gauge pressures add to a much smaller pressure, producing a lower-intensity sound. Although completely destructive interference is possible only under the simplest conditions, it is possible to reduce noise levels by 30 dB or more using this technique.

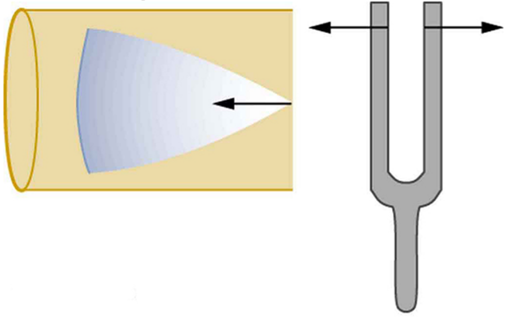

Noise Cancellation System
Noise sensor
Driver
Pressure servo
Cushions
Boom and cable
On/off switch
Power input
Boom mic with low distortion housing
Boom adjustment

Headphones designed to cancel noise with destructive interference create a sound wave exactly opposite to the incoming sound. These headphones can be more effective than the simple passive attenuation used in most ear protection. Such headphones were used on the record-setting, around the world nonstop flight of the Voyager aircraft to protect the pilots’ hearing from engine noise.

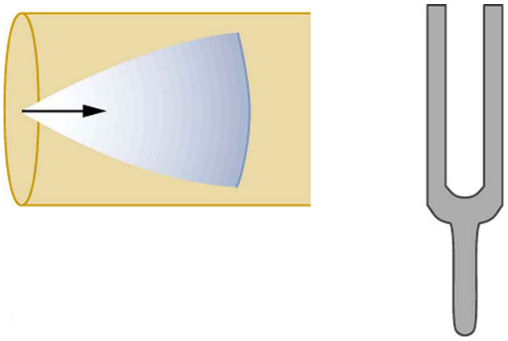
Where else can we observe sound interference? All sound resonances, such as in musical instruments, are due to constructive and destructive interference. Only the resonant frequencies interfere constructively to form standing waves, while others interfere destructively and are absent. From the toot made by blowing over a bottle, to the characteristic flavor of a violin’s sounding box, to the recognizability of a great singer’s voice, resonance and standing waves play a vital role.

Note:
Interference
Interference is such a fundamental aspect of waves that observing interference is proof that something is a wave. The wave nature of light was established by experiments showing interference. Similarly, when electrons scattered from crystals exhibited interference, their wave nature was confirmed to be exactly as predicted by symmetry with certain wave characteristics of light.

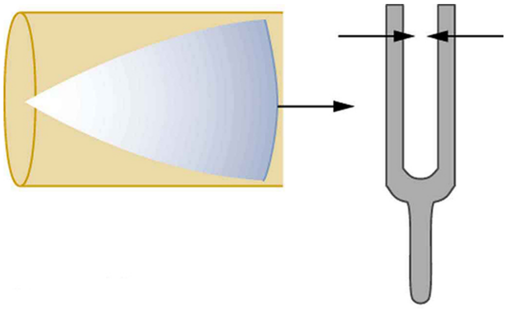
Suppose we hold a tuning fork near the end of a tube that is closed at the other end, as shown in [link], [link], [link], and [link]. If the tuning fork has just the right frequency, the air column in the tube resonates loudly, but at most frequencies it vibrates very little. This observation just means that the air column has only certain natural frequencies. The figures show how a resonance at the lowest of these natural frequencies is formed. A disturbance travels down the tube at the speed of sound and bounces off the closed end. If the tube is just the right length, the reflected sound arrives back at the tuning fork exactly half a cycle later, and it interferes constructively with the continuing sound produced by the tuning fork. The incoming and reflected sounds form a standing wave in the tube as shown.



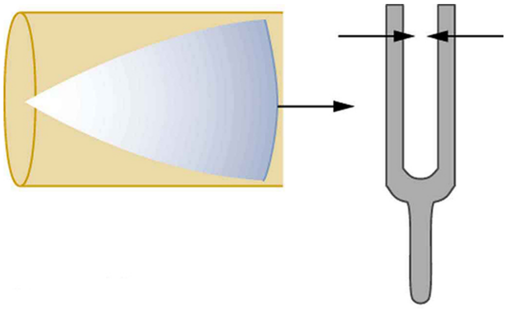
Resonance of air in a tube closed at one end, caused by a tuning fork. A disturbance moves down the tube.



Resonance of air in a tube closed at one end, caused by a tuning fork. The disturbance reflects from the closed end of the tube.

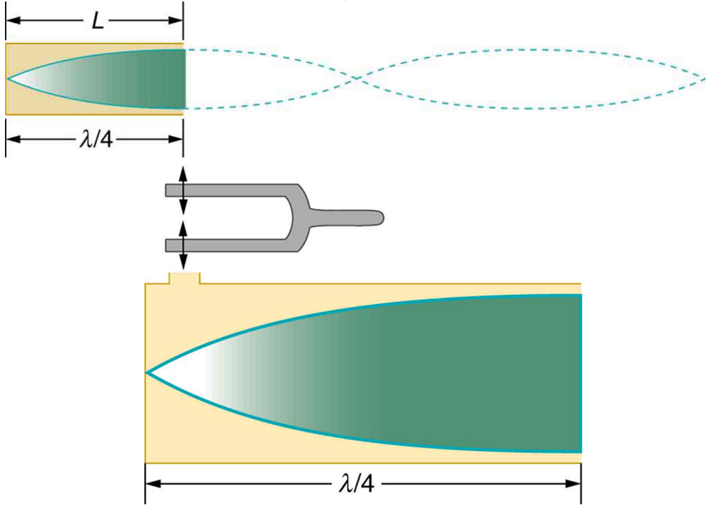


Resonance of air in a tube closed at one end, caused by a tuning fork. If the length of the tube L is just right, the disturbance gets back to the tuning fork half a cycle later and interferes constructively with the continuing sound from the tuning fork. This interference forms a standing wave, and the air column resonates.



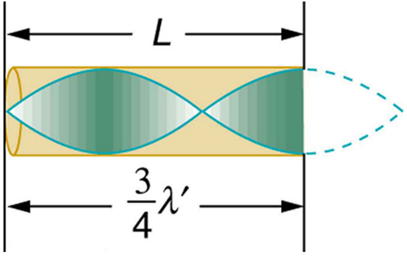
Resonance of air in a tube closed at one end, caused by a tuning fork. A graph of air displacement along the length of the tube shows none at the closed end, where the motion is constrained, and a maximum at the open end. This standing wave has one-fourth of its wavelength in the tube, so that $\lambda = 4L$.

The standing wave formed in the tube has its maximum air displacement (an **antinode**) at the open end, where motion is unconstrained, and no displacement (a **node**) at the closed end, where air movement is halted. The distance from a node to an antinode is one-fourth of a wavelength, and this equals the length of the tube; thus, $\lambda = 4L$. This same resonance can be produced by a vibration introduced at or near the closed end of the tube, as shown in [link]. It is best to consider this a natural vibration of the air column independently of how it is induced.

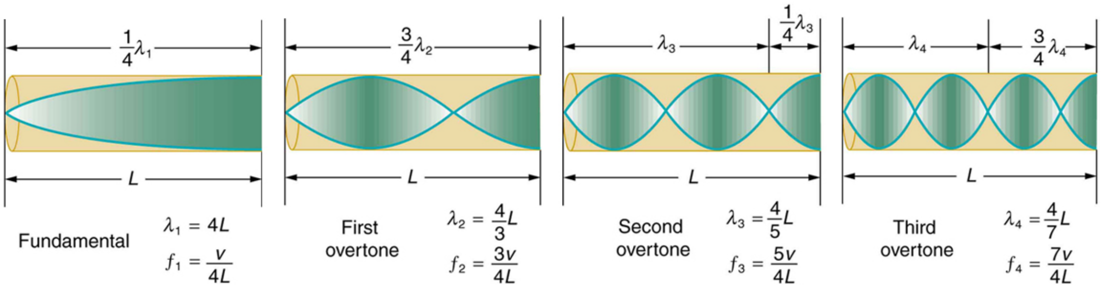


The same standing wave is created in the tube by a vibration introduced near its closed end.

Given that maximum air displacements are possible at the open end and none at the closed end, there are other, shorter wavelengths that can resonate in the tube, such as the one shown in [link]. Here the standing wave has three-fourths of its wavelength in the tube, or $L = (3/4)\lambda'$, so that $\lambda' = 4L/3$. Continuing this process reveals a whole series of shorter-wavelength and higher-frequency sounds that resonate in the tube. We use specific terms for the resonances in any system. The lowest resonant frequency is called the **fundamental**, while all higher resonant frequencies are called **overtones**. All resonant frequencies are integral multiples of the fundamental, and they are collectively called **harmonics**. The fundamental is the first harmonic, the first overtone is the second harmonic, and so on. [link] shows the fundamental and the first three overtones (the first four harmonics) in a tube closed at one end.

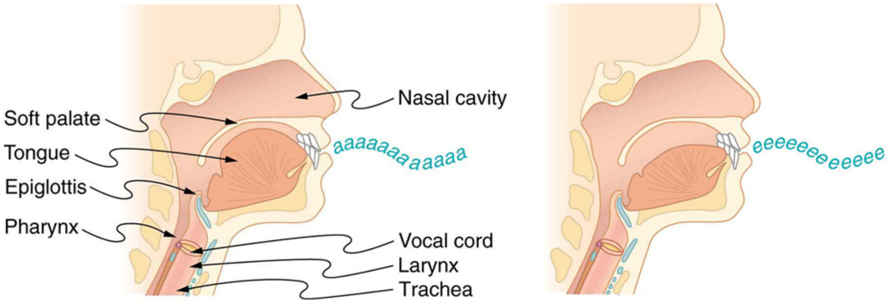


Another resonance for a tube closed at one end. This has maximum air displacements at the open end, and none at the closed end. The wavelength is shorter, with three-fourths λ' equaling the length of the tube, so that $\lambda' = 4L/3$. This higher-frequency vibration is the first overtone.


$\frac{1}{4}\lambda_1$
L
Fundamental
$\lambda_1 = 4L$
$f_1 = \frac{v}{4L}$
$\frac{3}{4}\lambda_2$
L
First overtone
$\lambda_2 = \frac{4}{3}L$
$f_2 = \frac{3v}{4L}$
$\frac{1}{4}\lambda_3$
λ_3
L
Second overtone
$\lambda_3 = \frac{4}{5}L$
$f_3 = \frac{5v}{4L}$
λ_4
$\frac{3}{4}\lambda_4$
L
Third overtone
$\lambda_4 = \frac{4}{7}L$
$f_4 = \frac{7v}{4L}$

The fundamental and three lowest overtones for a tube closed at one end. All have maximum air displacements at the open end and none at the closed end.

The fundamental and overtones can be present simultaneously in a variety of combinations. For example, middle C on a trumpet has a sound distinctively different from middle C on a clarinet, both instruments being modified versions of a tube closed at one end. The fundamental frequency is the same (and usually the most intense), but the overtones and their mix of intensities are different and subject to shading by the musician. This mix is what gives various musical instruments (and human voices) their distinctive characteristics, whether they have air columns, strings, sounding boxes, or drumheads. In fact, much of our speech is determined by shaping the cavity formed by the throat and mouth and positioning the tongue to adjust the fundamental and combination of overtones. Simple resonant cavities can be made to resonate with the sound of the vowels, for example. (See [link].) In boys, at puberty, the larynx grows and the shape of the resonant cavity changes giving rise to the difference in predominant frequencies in speech between men and women.


Soft palate
Tongue
Epiglottis
Pharynx
Nasal cavity
aaaaaaaaaaaa
Vocal cord
Larynx
Trachea
eeeeeeeeeeee

The throat and mouth form an air column closed at one end that resonates in response to vibrations in the voice box. The spectrum of overtones and their intensities vary with mouth shaping and tongue position to form different sounds. The voice box can be replaced with a mechanical vibrator, and understandable speech is still possible. Variations in basic shapes make different voices recognizable.

Now let us look for a pattern in the resonant frequencies for a simple tube that is closed at one end. The fundamental has $\lambda = 4L$, and frequency is related to wavelength and the speed of sound as given by:

Equation:

$$v_{\text{w}} = f\lambda.$$

Solving for f in this equation gives

Equation:

$$f = \frac{v_{\text{w}}}{\lambda} = \frac{v_{\text{w}}}{4L},$$

where v_{w} is the speed of sound in air. Similarly, the first overtone has $\lambda' = 4L/3$ (see [link]), so that

Equation:

$$f' = 3\frac{v_{\text{w}}}{4L} = 3f.$$

Because $f' = 3f$, we call the first overtone the third harmonic. Continuing this process, we see a pattern that can be generalized in a single expression. The resonant frequencies of a tube closed at one end are

Equation:

$$f_n = n\frac{v_\mathrm{w}}{4L}, n = 1,3,5,$$

where f_1 is the fundamental, f_3 is the first overtone, and so on. It is interesting that the resonant frequencies depend on the speed of sound and, hence, on temperature. This dependence poses a noticeable problem for organs in old unheated cathedrals, and it is also the reason why musicians commonly bring their wind instruments to room temperature before playing them.

Example:

Find the Length of a Tube with a 128 Hz Fundamental

(a) What length should a tube closed at one end have on a day when the air temperature, is $22.0^\circ\mathrm{C}$, if its fundamental frequency is to be 128 Hz (C below middle C)?

(b) What is the frequency of its fourth overtone?

Strategy

The length L can be found from the relationship in $f_n = n\frac{v_\mathrm{w}}{4L}$, but we will first need to find the speed of sound v_w.

Solution for (a)

(1) Identify knowns:

- the fundamental frequency is 128 Hz
- the air temperature is $22.0^\circ\mathrm{C}$

(2) Use $f_n = n\frac{v_\mathrm{w}}{4L}$ to find the fundamental frequency ($n = 1$).

Equation:

$$f_1 = \frac{v_\mathrm{w}}{4L}$$

(3) Solve this equation for length.

Equation:

$$L = \frac{v_\mathrm{w}}{4f_1}$$

(4) Find the speed of sound using $v_{\mathrm{w}} = (331\ \mathrm{m/s})\sqrt{\frac{T}{273\ \mathrm{K}}}$.

Equation:

$$v_{\mathrm{w}} = (331\ \mathrm{m/s})\sqrt{\frac{295\ \mathrm{K}}{273\ \mathrm{K}}} = 344\ \mathrm{m/s}$$

(5) Enter the values of the speed of sound and frequency into the expression for L.

Equation:

$$L = \frac{v_{\mathrm{w}}}{4f_1} = \frac{344\ \mathrm{m/s}}{4(128\ \mathrm{Hz})} = 0.672\ \mathrm{m}$$

Discussion on (a)

Many wind instruments are modified tubes that have finger holes, valves, and other devices for changing the length of the resonating air column and hence, the frequency of the note played. Horns producing very low frequencies, such as tubas, require tubes so long that they are coiled into loops.

Solution for (b)

(1) Identify knowns:

- the first overtone has $n = 3$
- the second overtone has $n = 5$
- the third overtone has $n = 7$
- the fourth overtone has $n = 9$

(2) Enter the value for the fourth overtone into $f_n = n\frac{v_{\mathrm{w}}}{4L}$.

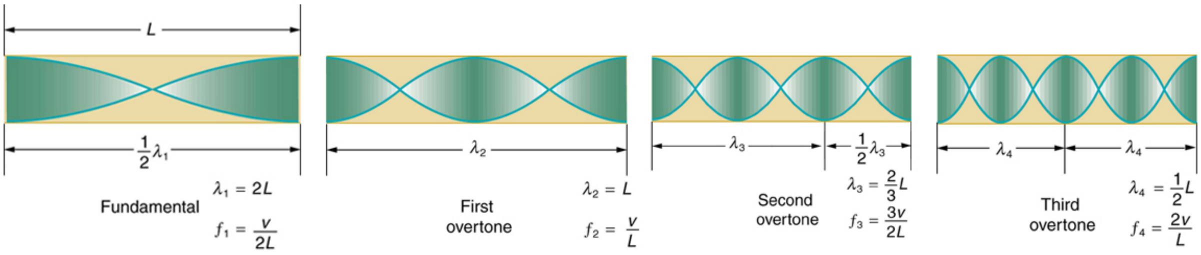
Equation:

$$f_9 = 9\frac{v_{\mathrm{w}}}{4L} = 9f_1 = 1.15\ \mathrm{kHz}$$

Discussion on (b)

Whether this overtone occurs in a simple tube or a musical instrument depends on how it is stimulated to vibrate and the details of its shape. The trombone, for example, does not produce its fundamental frequency and only makes overtones.

Another type of tube is one that is *open* at both ends. Examples are some organ pipes, flutes, and oboes. The resonances of tubes open at both ends can be analyzed in a very similar fashion to those for tubes closed at one end. The air columns in tubes open at both ends have maximum air displacements at both ends, as illustrated in [link]. Standing waves form as shown.


L
$\frac{1}{2}\lambda_1$
$\lambda_1 = 2L$
Fundamental
$f_1 = \frac{v}{2L}$
λ_2
First overtone
$\lambda_2 = L$
$f_2 = \frac{v}{L}$
λ_3
$\frac{1}{2}\lambda_3$
Second overtone
$\lambda_3 = \frac{2}{3}L$
$f_3 = \frac{3v}{2L}$
λ_4
λ_4
Third overtone
$\lambda_4 = \frac{1}{2}L$
$f_4 = \frac{2v}{L}$

The resonant frequencies of a tube open at both ends are shown, including the fundamental and the first three overtones. In all cases the maximum air displacements occur at both ends of the tube, giving it different natural frequencies than a tube closed at one end.

Based on the fact that a tube open at both ends has maximum air displacements at both ends, and using [link] as a guide, we can see that the resonant frequencies of a tube open at both ends are:

Equation:

$$f_n = n\frac{v_\text{w}}{2L},\ n = 1, 2, 3...,$$

where f_1 is the fundamental, f_2 is the first overtone, f_3 is the second overtone, and so on. Note that a tube open at both ends has a fundamental frequency twice what it would have if closed at one end. It also has a different spectrum of overtones than a tube closed at one end. So if you had two tubes with the same fundamental frequency but one was open at both ends and the other was closed at one end, they would sound different when played because they have different overtones. Middle C, for example, would sound richer played on an open tube, because it has even multiples of the fundamental as well as odd. A closed tube has only odd multiples.

Note:

Real-World Applications: Resonance in Everyday Systems

Resonance occurs in many different systems, including strings, air columns, and atoms. Resonance is the driven or forced oscillation of a system at its natural frequency. At resonance, energy is transferred rapidly to the oscillating system, and the amplitude of its oscillations grows until the system can no longer be described by Hooke’s law. An example of this is the distorted sound intentionally produced in certain types of rock music.

Wind instruments use resonance in air columns to amplify tones made by lips or vibrating reeds. Other instruments also use air resonance in clever ways to amplify sound. [link] shows a violin and a guitar, both of which have sounding boxes but with different shapes, resulting in different overtone structures. The vibrating string creates a sound that resonates in the sounding box, greatly amplifying the sound and creating overtones that give the instrument its characteristic flavor. The more complex the shape of the sounding box, the greater its ability to resonate over a wide range of frequencies. The marimba, like the one shown in [link] uses pots or gourds below the wooden slats to amplify their tones. The resonance of the pot can be adjusted by adding water.

String instruments such as violins and guitars use resonance in their sounding boxes to amplify and enrich the sound created by their vibrating strings. The bridge and supports couple the string vibrations to the sounding boxes and air within. (credits: guitar, Feliciano Guimares, Fotopedia; violin, Steve Snodgrass, Flickr)

Resonance has been used in musical instruments since prehistoric times. This marimba uses gourds as resonance chambers to amplify its sound. (credit: APC Events, Flickr)

We have emphasized sound applications in our discussions of resonance and standing waves, but these ideas apply to any system that has wave characteristics. Vibrating strings, for example, are actually resonating and have fundamentals and overtones similar to those for air columns. More subtle are the resonances in atoms due to the wave character of their electrons. Their orbitals can be viewed as standing waves, which have a fundamental (ground state) and overtones (excited states). It is fascinating that wave characteristics apply to such a wide range of physical systems.

Exercise:

Check Your Understanding

Problem:

Describe how noise-canceling headphones differ from standard headphones used to block outside sounds.

Solution:

Regular headphones only block sound waves with a physical barrier. Noise-canceling headphones use destructive interference to reduce the loudness of outside sounds.

Exercise:
Check Your Understanding

Problem:

How is it possible to use a standing wave's node and antinode to determine the length of a closed-end tube?

Solution:

When the tube resonates at its natural frequency, the wave's node is located at the closed end of the tube, and the antinode is located at the open end. The length of the tube is equal to one-fourth of the wavelength of this wave. Thus, if we know the wavelength of the wave, we can determine the length of the tube.

Note:
PhET Explorations: Sound
This simulation lets you see sound waves. Adjust the frequency or volume and you can see and hear how the wave changes. Move the listener around and hear what she hears.
https://archive.cnx.org/specials/c4d3b96e-41f3-11e5-ab7b-47e22dffc18e/sound/#sim-single-source

Section Summary

- Sound interference and resonance have the same properties as defined for all waves.
- In air columns, the lowest-frequency resonance is called the fundamental, whereas all higher resonant frequencies are called overtones. Collectively, they are called harmonics.

- The resonant frequencies of a tube closed at one end are:
  Equation:

  $$f_n = n\frac{v_\text{w}}{4L}, n = 1, 3, 5...,$$

  f_1 is the fundamental and L is the length of the tube.
- The resonant frequencies of a tube open at both ends are:
  Equation:

  $$f_n = n\frac{v_\text{w}}{2L}, n = 1, 2, 3...$$

Conceptual Questions

Exercise:

Problem:

How does an unamplified guitar produce sounds so much more intense than those of a plucked string held taut by a simple stick?

Exercise:

Problem:

You are given two wind instruments of identical length. One is open at both ends, whereas the other is closed at one end. Which is able to produce the lowest frequency?

Exercise:

Problem:

What is the difference between an overtone and a harmonic? Are all harmonics overtones? Are all overtones harmonics?

Problems & Exercises

Exercise:

Problem:

A “showy” custom-built car has two brass horns that are supposed to produce the same frequency but actually emit 263.8 and 264.5 Hz. What beat frequency is produced?

Solution:

0.7 Hz

Exercise:

Problem:

What beat frequencies will be present: (a) If the musical notes A and C are played together (frequencies of 220 and 264 Hz)? (b) If D and F are played together (frequencies of 297 and 352 Hz)? (c) If all four are played together?

Exercise:

Problem:

What beat frequencies result if a piano hammer hits three strings that emit frequencies of 127.8, 128.1, and 128.3 Hz?

Solution:

0.3 Hz, 0.2 Hz, 0.5 Hz

Exercise:

Problem:

A piano tuner hears a beat every 2.00 s when listening to a 264.0-Hz tuning fork and a single piano string. What are the two possible frequencies of the string?

Exercise:

Problem:

(a) What is the fundamental frequency of a 0.672-m-long tube, open at both ends, on a day when the speed of sound is 344 m/s? (b) What is the frequency of its second harmonic?

Solution:

(a) 256 Hz

(b) 512 Hz

Exercise:

Problem:

If a wind instrument, such as a tuba, has a fundamental frequency of 32.0 Hz, what are its first three overtones? It is closed at one end. (The overtones of a real tuba are more complex than this example, because it is a tapered tube.)

Exercise:

Problem:

What are the first three overtones of a bassoon that has a fundamental frequency of 90.0 Hz? It is open at both ends. (The overtones of a real bassoon are more complex than this example, because its double reed makes it act more like a tube closed at one end.)

Solution:

180 Hz, 270 Hz, 360 Hz

Exercise:

Problem:

How long must a flute be in order to have a fundamental frequency of 262 Hz (this frequency corresponds to middle C on the evenly tempered chromatic scale) on a day when air temperature is 20.0°C? It is open at both ends.

Exercise:

Problem:

What length should an oboe have to produce a fundamental frequency of 110 Hz on a day when the speed of sound is 343 m/s? It is open at both ends.

Solution:

1.56 m

Exercise:

Problem:

What is the length of a tube that has a fundamental frequency of 176 Hz and a first overtone of 352 Hz if the speed of sound is 343 m/s?

Exercise:

Problem:

(a) Find the length of an organ pipe closed at one end that produces a fundamental frequency of 256 Hz when air temperature is 18.0°C. (b) What is its fundamental frequency at 25.0°C?

Solution:

(a) 0.334 m

(b) 259 Hz

Exercise:

Problem:

By what fraction will the frequencies produced by a wind instrument change when air temperature goes from 10.0°C to 30.0°C? That is, find the ratio of the frequencies at those temperatures.

Exercise:

Problem:

The ear canal resonates like a tube closed at one end. (See [link].) If ear canals range in length from 1.80 to 2.60 cm in an average population, what is the range of fundamental resonant frequencies? Take air temperature to be 37.0°C, which is the same as body temperature. How does this result correlate with the intensity versus frequency graph ([link] of the human ear?

Solution:

3.39 to 4.90 kHz

Exercise:

Problem:

Calculate the first overtone in an ear canal, which resonates like a 2.40-cm-long tube closed at one end, by taking air temperature to be 37.0°C. Is the ear particularly sensitive to such a frequency? (The resonances of the ear canal are complicated by its nonuniform shape, which we shall ignore.)

Exercise:

Problem:

A crude approximation of voice production is to consider the breathing passages and mouth to be a resonating tube closed at one end. (See [link].) (a) What is the fundamental frequency if the tube is 0.240-m long, by taking air temperature to be 37.0°C? (b) What would this frequency become if the person replaced the air with helium? Assume the same temperature dependence for helium as for air.

Solution:

(a) 367 Hz

(b) 1.07 kHz

Exercise:

Problem:

(a) Students in a physics lab are asked to find the length of an air column in a tube closed at one end that has a fundamental frequency of 256 Hz. They hold the tube vertically and fill it with water to the top, then lower the water while a 256-Hz tuning fork is rung and listen for the first resonance. What is the air temperature if the resonance occurs for a length of 0.336 m? (b) At what length will they observe the second resonance (first overtone)?

Exercise:

Problem:

What frequencies will a 1.80-m-long tube produce in the audible range at 20.0°C if: (a) The tube is closed at one end? (b) It is open at both ends?

Solution:

(a) $f_n = n(47.6\text{ Hz}),\ n = 1, 3, 5, \ldots, 419$

(b) $f_n = n(95.3\text{ Hz}),\ n = 1, 2, 3, \ldots, 210$

Glossary

antinode
: point of maximum displacement

node

point of zero displacement

fundamental
the lowest-frequency resonance

overtones
all resonant frequencies higher than the fundamental

harmonics
the term used to refer collectively to the fundamental and its overtones

Hearing

- Define hearing, pitch, loudness, timbre, note, tone, phon, ultrasound, and infrasound.
- Compare loudness to frequency and intensity of a sound.
- Identify structures of the inner ear and explain how they relate to sound perception.

Hearing allows this vocalist, his band, and his fans to enjoy music. (credit: West Point Public Affairs, Flickr)

The human ear has a tremendous range and sensitivity. It can give us a wealth of simple information—such as pitch, loudness, and direction. And from its input we can detect musical quality and nuances of voiced emotion. How is our hearing related to the physical qualities of sound, and how does the hearing mechanism work?

Hearing is the perception of sound. (Perception is commonly defined to be awareness through the senses, a typically circular definition of higher-level processes in living organisms.) Normal human hearing encompasses frequencies from 20 to 20,000 Hz, an impressive range. Sounds below 20 Hz are called **infrasound**, whereas those above 20,000 Hz are **ultrasound**. Neither is perceived by the ear, although infrasound can sometimes be felt as vibrations. When we do hear low-frequency vibrations, such as the sounds of a diving board, we hear the individual vibrations only because there are higher-frequency sounds in each. Other animals have hearing ranges different from that of humans. Dogs can hear sounds as high as 30,000 Hz, whereas bats and dolphins can hear up to 100,000-Hz sounds. You may have noticed that dogs respond to the sound of a dog whistle which produces sound out of the range of human hearing. Elephants are known to respond to frequencies below 20 Hz.

The perception of frequency is called **pitch**. Most of us have excellent relative pitch, which means that we can tell whether one sound has a different frequency from another. Typically, we can discriminate between two sounds if their frequencies differ by 0.3% or more. For example, 500.0 and 501.5 Hz are noticeably different. Pitch perception is directly related to frequency and is not greatly affected by other physical quantities such as intensity. Musical **notes** are particular sounds that can be produced by most instruments and in Western music have particular names. Combinations of notes constitute music. Some people can identify musical notes, such as A-sharp, C, or E-flat, just by listening to them. This uncommon ability is called perfect pitch.

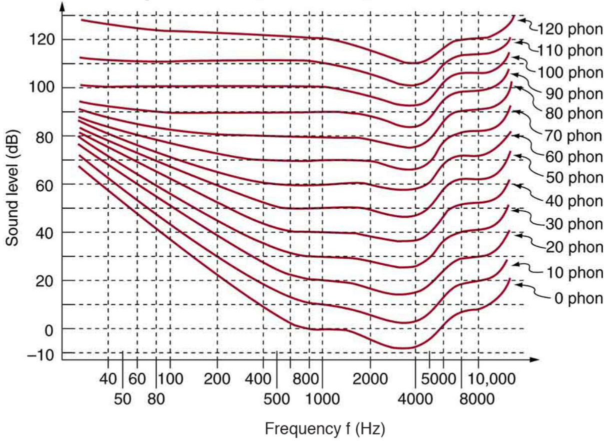
The ear is remarkably sensitive to low-intensity sounds. The lowest audible intensity or threshold is about $10^{-12}\ \text{W/m}^2$ or 0 dB. Sounds as much as 10^{12} more intense can be briefly tolerated. Very few measuring devices are capable of observations over a range of a trillion. The perception of intensity is called **loudness**. At a given frequency, it is possible to discern differences of about 1 dB, and a change of 3 dB is easily noticed. But loudness is not related to intensity alone. Frequency has a major effect on how loud a sound seems. The ear has its maximum sensitivity to frequencies in the range of 2000 to 5000 Hz, so that sounds in this range are perceived as being louder than, say, those at 500 or 10,000 Hz, even when they all have the same intensity. Sounds near the high- and low-frequency extremes of the hearing range seem even less loud, because the ear is even less sensitive at those frequencies. [link] gives the dependence of certain human hearing perceptions on physical quantities.

Perception	**Physical quantity**
Pitch	Frequency
Loudness	Intensity and Frequency
Timbre	Number and relative intensity of multiple frequencies. Subtle craftsmanship leads to non-linear effects and more detail.
Note	Basic unit of music with specific names, combined to generate tunes
Tone	Number and relative intensity of multiple frequencies.

Sound Perceptions

When a violin plays middle C, there is no mistaking it for a piano playing the same note. The reason is that each instrument produces a distinctive set of frequencies and intensities. We call our perception of these combinations of frequencies and intensities **tone** quality, or more commonly the **timbre** of the sound. It is more difficult to correlate timbre perception to physical quantities than it is for loudness or pitch perception. Timbre is more subjective. Terms such as dull, brilliant, warm, cold, pure, and rich are employed to describe the timbre of a sound. So the consideration of timbre takes us into the realm of perceptual psychology, where higher-level processes in the brain are dominant. This is true for other perceptions of sound, such as music and noise. We shall not delve further into them; rather, we will concentrate on the question of loudness perception.

A unit called a **phon** is used to express loudness numerically. Phons differ from decibels because the phon is a unit of loudness perception, whereas the decibel is a unit of physical intensity. [link] shows the relationship of loudness to intensity (or intensity level) and frequency for persons with normal hearing. The curved lines are equal-loudness curves. Each curve is labeled with its loudness in phons. Any sound along a given curve will be perceived as equally loud by the average person. The curves were determined by having large numbers of people compare the loudness of sounds at different frequencies and sound intensity levels. At a frequency of 1000 Hz, phons are taken to be numerically equal to decibels. The following example helps illustrate how to use the graph:



The relationship of loudness in phons to intensity level (in decibels) and intensity (in watts per meter squared) for persons with normal hearing. The curved lines are equal-loudness curves—all sounds on a given curve are perceived as equally loud. Phons and decibels are defined to be the same at 1000 Hz.

Example:

Measuring Loudness: Loudness Versus Intensity Level and Frequency

(a) What is the loudness in phons of a 100-Hz sound that has an intensity level of 80 dB? (b) What is the intensity level in decibels of a 4000-Hz sound having a loudness of 70 phons? (c) At what intensity level will an 8000-Hz sound have the same loudness as a 200-Hz sound at 60 dB?

Strategy for (a)

The graph in [link] should be referenced in order to solve this example. To find the loudness of a given sound, you must know its frequency and intensity level and locate that point on the square grid, then interpolate between loudness curves to get the loudness in phons.

Solution for (a)

(1) Identify knowns:

- The square grid of the graph relating phons and decibels is a plot of intensity level versus frequency—both physical quantities.
- 100 Hz at 80 dB lies halfway between the curves marked 70 and 80 phons.

(2) Find the loudness: 75 phons.

Strategy for (b)

The graph in [link] should be referenced in order to solve this example. To find the intensity level of a sound, you must have its frequency and loudness. Once that point is located, the intensity level can be determined from the vertical axis.

Solution for (b)

(1) Identify knowns:

- Values are given to be 4000 Hz at 70 phons.

(2) Follow the 70-phon curve until it reaches 4000 Hz. At that point, it is below the 70 dB line at about 67 dB.

(3) Find the intensity level:

67 dB

Strategy for (c)

The graph in [link] should be referenced in order to solve this example.

Solution for (c)

(1) Locate the point for a 200 Hz and 60 dB sound.

(2) Find the loudness: This point lies just slightly above the 50-phon curve, and so its loudness is 51 phons.

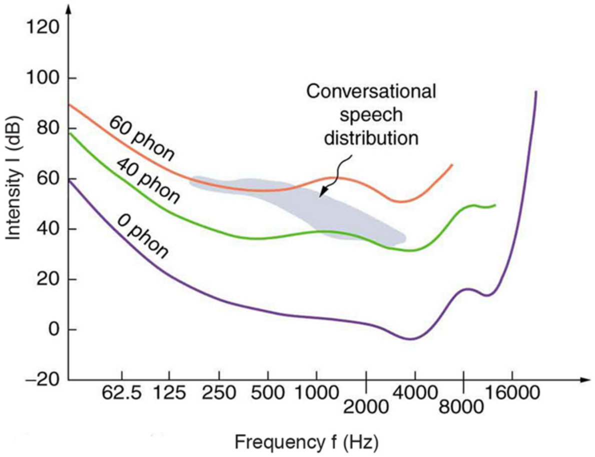
(3) Look for the 51-phon level is at 8000 Hz: 63 dB.

Discussion

These answers, like all information extracted from [link], have uncertainties of several phons or several decibels, partly due to difficulties in interpolation, but mostly related to uncertainties in the equal-loudness curves.

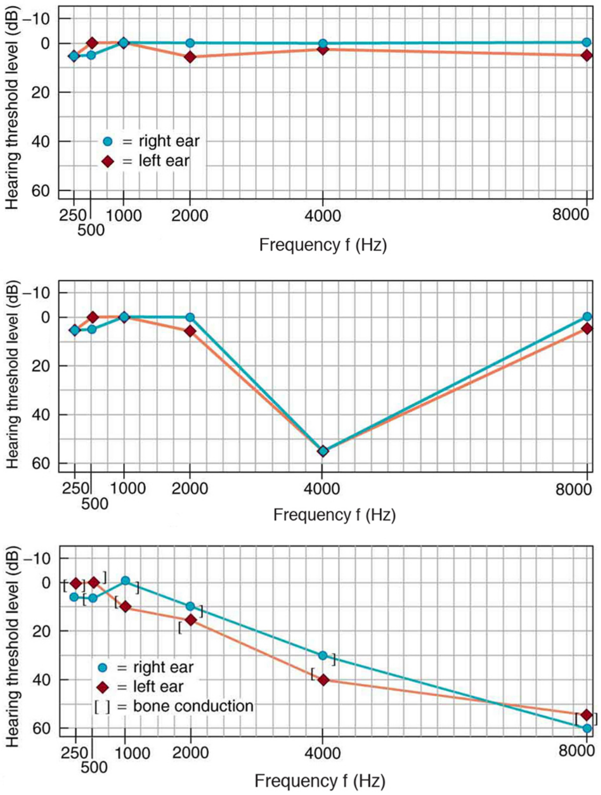
Further examination of the graph in [link] reveals some interesting facts about human hearing. First, sounds below the 0-phon curve are not perceived by most people. So, for example, a 60 Hz sound at 40 dB is inaudible. The 0-phon curve represents the threshold of normal hearing. We can hear some sounds at intensity levels below 0 dB. For example, a 3-dB, 5000-Hz sound is audible, because it lies above the 0-phon curve. The loudness curves all have dips in them between about 2000 and 5000 Hz. These dips mean the ear is most sensitive to frequencies in that range. For example, a 15-dB sound at 4000 Hz has a loudness of 20 phons, the same as a 20-dB sound at 1000 Hz. The curves rise at both extremes of the frequency range, indicating that a greater-intensity level sound is needed at those frequencies to be perceived to be as loud as at middle frequencies. For example, a sound at 10,000 Hz must have an intensity level of 30 dB to seem as loud as a 20 dB sound at 1000 Hz. Sounds above 120 phons are painful as well as damaging.

We do not often utilize our full range of hearing. This is particularly true for frequencies above 8000 Hz, which are rare in the environment and are unnecessary for understanding conversation or appreciating music. In fact, people who have lost the ability to hear such high frequencies are usually unaware of their loss until tested. The shaded region in [link] is the frequency and intensity region where most conversational sounds fall. The curved lines indicate what effect hearing losses of 40 and 60 phons will have. A 40-phon hearing loss at all frequencies still allows a person to understand conversation, although it will seem very quiet. A person with a 60-phon loss at all frequencies will hear only the lowest frequencies and will not be able to understand speech unless it is much louder than normal. Even so, speech may seem indistinct, because higher frequencies are not as well perceived. The conversational speech region also has a gender component, in that female voices are usually characterized by higher frequencies. So the person with a 60-phon hearing impediment might have difficulty understanding the normal conversation of a woman.



The shaded region represents frequencies and intensity levels found in normal conversational speech. The 0-phon line represents the normal hearing threshold, while those at 40 and 60 represent thresholds for people with 40- and 60-phon hearing losses, respectively.

Hearing tests are performed over a range of frequencies, usually from 250 to 8000 Hz, and can be displayed graphically in an audiogram like that in [link]. The hearing threshold is measured in dB *relative to the normal threshold*, so that normal hearing registers as 0 dB at all frequencies. Hearing loss caused by noise typically shows a dip near the 4000 Hz frequency, irrespective of the frequency that caused the loss and often affects both ears. The most common form of hearing loss comes with age and is called *presbycusis*—literally elder ear. Such loss is increasingly severe at higher frequencies, and interferes with music appreciation and speech recognition.

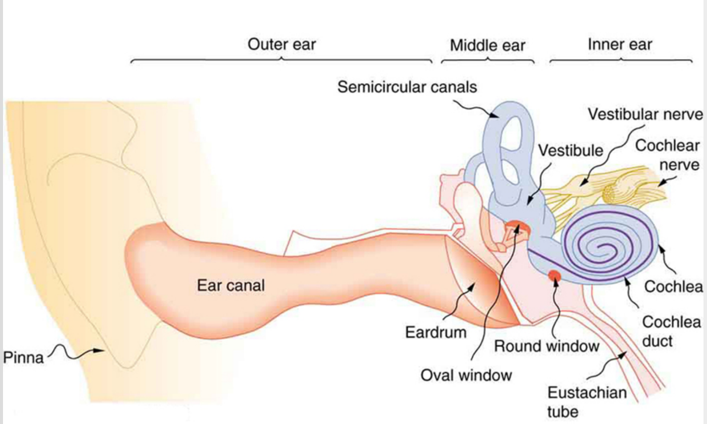


Audiograms showing the threshold in intensity level versus frequency for three different individuals. Intensity level is measured relative to the normal threshold. The top left graph is that of a person with normal hearing. The graph to its right has a dip at 4000 Hz and is that of a child who suffered hearing loss due to a cap gun. The third graph is typical of presbycusis, the progressive loss of higher frequency hearing with age. Tests performed by bone conduction (brackets) can distinguish nerve damage from middle ear damage.

Note:

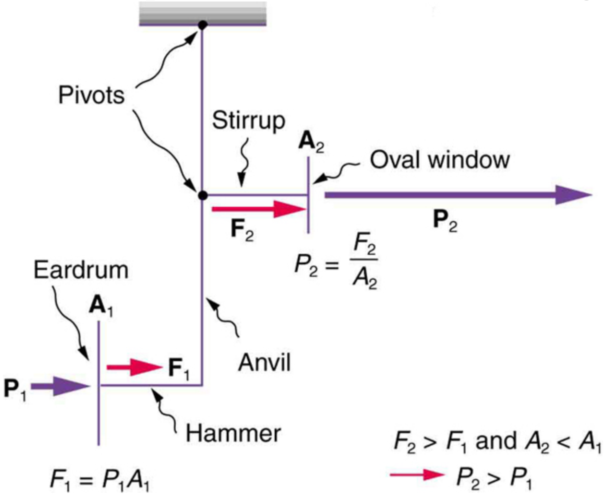
The Hearing Mechanism

The hearing mechanism involves some interesting physics. The sound wave that impinges upon our ear is a pressure wave. The ear is a transducer that converts sound waves into electrical nerve impulses in a manner much more sophisticated than, but analogous to, a microphone. [link] shows the gross anatomy of the ear with its division into three parts: the outer ear or ear canal; the middle ear, which runs from the eardrum to the cochlea; and the inner ear, which is the cochlea itself. The body part normally referred to as the ear is technically called the pinna.


Outer ear
Middle ear
Inner ear
Semicircular canals
Vestibular nerve
Vestibule
Cochlear nerve
Ear canal
Cochlea
Cochlea duct
Eardrum
Round window
Pinna
Oval window
Eustachian tube

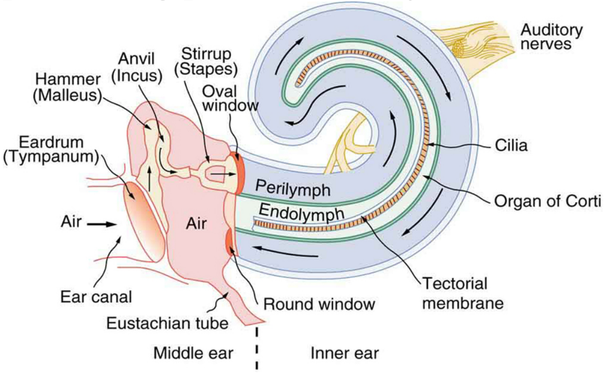
The illustration shows the gross anatomy of the human ear.

The outer ear, or ear canal, carries sound to the recessed protected eardrum. The air column in the ear canal resonates and is partially responsible for the sensitivity of the ear to sounds in the 2000 to 5000 Hz range. The middle ear converts sound into mechanical vibrations and applies these vibrations to the cochlea. The lever system of the middle ear takes the force exerted on the eardrum by sound pressure variations, amplifies it and transmits it to the inner ear via the oval window, creating pressure waves in the cochlea approximately 40 times greater than those impinging on the eardrum. (See [link].) Two muscles in the middle ear (not shown) protect the inner ear from very intense sounds. They react to intense sound in a few milliseconds and reduce the force transmitted to the cochlea. This protective reaction can also be triggered by your own voice, so that humming while shooting a gun, for example, can reduce noise damage.


Pivots
Stirrup
A_2
Oval window
F_2
P_2
$P_2 = \frac{F_2}{A_2}$
Eardrum
A_1
F_1
Anvil
P_1
Hammer
$F_2 > F_1$ and $A_2 < A_1$
$\longrightarrow P_2 > P_1$
$F_1 = P_1A_1$

This schematic shows the middle ear’s system for converting sound pressure into force, increasing that force through a lever system, and applying the increased force to a small area of the cochlea, thereby creating a pressure about 40 times that in the original sound wave. A protective muscle reaction to intense sounds greatly reduces the mechanical advantage of the lever system.

[link] shows the middle and inner ear in greater detail. Pressure waves moving through the cochlea cause the tectorial membrane to vibrate, rubbing cilia (called hair cells), which stimulate nerves that send electrical signals to the brain. The membrane resonates at different positions for different frequencies, with high frequencies stimulating nerves at the near end and low frequencies at the far end. The complete operation of the cochlea is still not understood, but several mechanisms for sending information to the brain are known to be involved. For sounds below about 1000 Hz, the nerves send signals at the same frequency as the sound. For frequencies greater than about 1000 Hz, the nerves signal frequency by position. There is a structure to the cilia, and there are connections between nerve cells that perform signal processing before information is sent to the brain. Intensity information is partly indicated by the number of nerve signals and by volleys of signals. The brain processes the cochlear nerve signals to provide additional information such as source direction (based on time and intensity comparisons of sounds from both ears). Higher-level processing produces many nuances, such as music appreciation.


Auditory nerves
Anvil (Incus)
Stirrup (Stapes)
Hammer (Malleus)
Oval window
Eardrum (Tympanum)
Cilia
Perilymph
Endolymph
Organ of Corti
Air
Air
Ear canal
Eustachian tube
Round window
Tectorial membrane
Middle ear
Inner ear

The inner ear, or cochlea, is a coiled tube about 3 mm in diameter and 3 cm in length if uncoiled. When the oval window is forced inward, as shown, a pressure wave travels through the perilymph in the direction of the arrows, stimulating nerves at the base of cilia in the organ of Corti.

Hearing losses can occur because of problems in the middle or inner ear. Conductive losses in the middle ear can be partially overcome by sending sound vibrations to the cochlea through the skull. Hearing aids for this purpose usually press against the bone behind the ear, rather than simply amplifying the sound sent into the ear canal as many hearing aids do. Damage to the nerves in the cochlea is not repairable, but amplification can partially compensate. There is a risk that amplification will produce further damage. Another common failure in the cochlea is damage or loss of the cilia but with nerves remaining functional. Cochlear implants that stimulate the nerves directly are now available and widely accepted. Over 100,000 implants are in use, in about equal numbers of adults and children.

The cochlear implant was pioneered in Melbourne, Australia, by Graeme Clark in the 1970s for his deaf father. The implant consists of three external components and two internal components. The external components are a microphone for picking up sound and converting it into an electrical signal, a speech processor to select certain frequencies and a transmitter to transfer the signal to the internal components through electromagnetic induction. The internal components consist of a receiver/transmitter secured in the bone beneath the skin, which converts the signals into electric impulses and sends them through an internal cable to the cochlea and an array of about 24 electrodes wound through the cochlea. These electrodes in turn send the impulses directly into the brain. The electrodes basically emulate the cilia.

Exercise:

Check Your Understanding

Problem:

Are ultrasound and infrasound imperceptible to all hearing organisms? Explain your answer.

Solution:

No, the range of perceptible sound is based in the range of human hearing. Many other organisms perceive either infrasound or ultrasound.

Section Summary

- The range of audible frequencies is 20 to 20,000 Hz.
- Those sounds above 20,000 Hz are ultrasound, whereas those below 20 Hz are infrasound.
- The perception of frequency is pitch.
- The perception of intensity is loudness.
- Loudness has units of phons.

Conceptual Questions

Exercise:

Problem:

Why can a hearing test show that your threshold of hearing is 0 dB at 250 Hz, when [link] implies that no one can hear such a frequency at less than 20 dB?

Problems & Exercises

Exercise:

Problem:

The factor of 10^{-12} in the range of intensities to which the ear can respond, from threshold to that causing damage after brief exposure, is truly remarkable. If you could measure distances over the same range with a single instrument and the smallest distance you could measure was 1 mm, what would the largest be?

Solution:
Equation:

$$1 \times 10^{6}\ \text{km}$$

Exercise:

Problem:

The frequencies to which the ear responds vary by a factor of 10^3. Suppose the speedometer on your car measured speeds differing by the same factor of 10^3, and the greatest speed it reads is 90.0 mi/h. What would be the slowest nonzero speed it could read?

Exercise:

Problem:

What are the closest frequencies to 500 Hz that an average person can clearly distinguish as being different in frequency from 500 Hz? The sounds are not present simultaneously.

Solution:

498.5 or 501.5 Hz

Exercise:

Problem:

Can the average person tell that a 2002-Hz sound has a different frequency than a 1999-Hz sound without playing them simultaneously?

Exercise:

Problem:

If your radio is producing an average sound intensity level of 85 dB, what is the next lowest sound intensity level that is clearly less intense?

Solution:

82 dB

Exercise:

Problem:

Can you tell that your roommate turned up the sound on the TV if its average sound intensity level goes from 70 to 73 dB?

Exercise:

Problem:

Based on the graph in [link], what is the threshold of hearing in decibels for frequencies of 60, 400, 1000, 4000, and 15,000 Hz? Note that many AC electrical appliances produce 60 Hz, music is commonly 400 Hz, a reference frequency is 1000 Hz, your maximum sensitivity is near 4000 Hz, and many older TVs produce a 15,750 Hz whine.

Solution:

approximately 48, 9, 0, –7, and 20 dB, respectively

Exercise:

Problem:

What sound intensity levels must sounds of frequencies 60, 3000, and 8000 Hz have in order to have the same loudness as a 40-dB sound of frequency 1000 Hz (that is, to have a loudness of 40 phons)?

Exercise:

Problem:

What is the approximate sound intensity level in decibels of a 600-Hz tone if it has a loudness of 20 phons? If it has a loudness of 70 phons?

Solution:

(a) 23 dB

(b) 70 dB

Exercise:

Problem:

(a) What are the loudnesses in phons of sounds having frequencies of 200, 1000, 5000, and 10,000 Hz, if they are all at the same 60.0-dB sound intensity level? (b) If they are all at 110 dB? (c) If they are all at 20.0 dB?

Exercise:

Problem:

Suppose a person has a 50-dB hearing loss at all frequencies. By how many factors of 10 will low-intensity sounds need to be amplified to seem normal to this person? Note that smaller amplification is appropriate for more intense sounds to avoid further hearing damage.

Solution:

Five factors of 10

Exercise:

Problem:

If a woman needs an amplification of 5.0×10^{12} times the threshold intensity to enable her to hear at all frequencies, what is her overall hearing loss in dB? Note that smaller amplification is appropriate for more intense sounds to avoid further damage to her hearing from levels above 90 dB.

Exercise:

Problem:

(a) What is the intensity in watts per meter squared of a just barely audible 200-Hz sound? (b) What is the intensity in watts per meter squared of a barely audible 4000-Hz sound?

Solution:

(a) $2 \times 10^{-10}\ \mathrm{W/m^2}$

(b) $2 \times 10^{-13}\ \mathrm{W/m^2}$

Exercise:

Problem:

(a) Find the intensity in watts per meter squared of a 60.0-Hz sound having a loudness of 60 phons. (b) Find the intensity in watts per meter squared of a 10,000-Hz sound having a loudness of 60 phons.

Exercise:

Problem:

A person has a hearing threshold 10 dB above normal at 100 Hz and 50 dB above normal at 4000 Hz. How much more intense must a 100-Hz tone be than a 4000-Hz tone if they are both barely audible to this person?

Solution:

2.5

Exercise:

Problem:

A child has a hearing loss of 60 dB near 5000 Hz, due to noise exposure, and normal hearing elsewhere. How much more intense is a 5000-Hz tone than a 400-Hz tone if they are both barely audible to the child?

Exercise:

Problem:

What is the ratio of intensities of two sounds of identical frequency if the first is just barely discernible as louder to a person than the second?

Solution:

1.26

Glossary

loudness
the perception of sound intensity

timbre
number and relative intensity of multiple sound frequencies

note
basic unit of music with specific names, combined to generate tunes

tone
number and relative intensity of multiple sound frequencies

phon
the numerical unit of loudness

ultrasound
sounds above 20,000 Hz

infrasound
sounds below 20 Hz

Ultrasound

- Define acoustic impedance and intensity reflection coefficient.
- Describe medical and other uses of ultrasound technology.
- Calculate acoustic impedance using density values and the speed of ultrasound.
- Calculate the velocity of a moving object using Doppler-shifted ultrasound.

ATL
Map 3
170dB/C 4
Persist Off
2D Opt:HSCT
Fr Rate:Surv
SonoCT®
XRes™
5
10
THUMB SUCKER

Ultrasound is used in medicine to painlessly and noninvasively monitor patient health and diagnose a wide range of disorders. (credit: abbybatchelder, Flickr)

Any sound with a frequency above 20,000 Hz (or 20 kHz)—that is, above the highest audible frequency—is defined to be ultrasound. In practice, it is possible to create ultrasound frequencies up to more than a gigahertz. (Higher frequencies are difficult to create; furthermore, they propagate poorly because they are very strongly absorbed.) Ultrasound has a tremendous number of applications, which range from burglar alarms to use in cleaning delicate objects to the guidance systems of bats. We begin our discussion of ultrasound with some of its applications in medicine, in which it is used extensively both for diagnosis and for therapy.

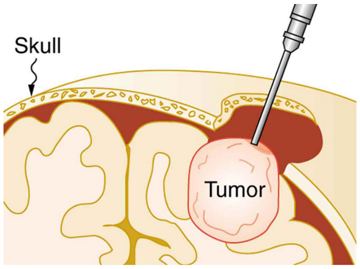
Note:

Characteristics of Ultrasound

The characteristics of ultrasound, such as frequency and intensity, are wave properties common to all types of waves. Ultrasound also has a wavelength that limits the fineness of detail it can detect. This characteristic is true of all waves. We can never observe details significantly smaller than the wavelength of our probe; for example, we will never see individual atoms with visible light, because the atoms are so small compared with the wavelength of light.

Ultrasound in Medical Therapy

Ultrasound, like any wave, carries energy that can be absorbed by the medium carrying it, producing effects that vary with intensity. When focused to intensities of 10^3 to $10^5\ \mathrm{W/m^2}$, ultrasound can be used to shatter gallstones or pulverize cancerous tissue in surgical procedures. (See [link].) Intensities this great can damage individual cells, variously causing their protoplasm to stream inside them, altering their permeability, or rupturing their walls through *cavitation*. Cavitation is the creation of vapor cavities in a fluid—the longitudinal vibrations in ultrasound alternatively compress and expand the medium, and at sufficient amplitudes the expansion separates molecules. Most cavitation damage is done when the cavities collapse, producing even greater shock pressures.



The tip of this small probe oscillates at 23 kHz with such a large amplitude that it pulverizes tissue on contact. The debris is then aspirated. The speed of the tip may exceed the speed of sound in tissue, thus creating shock waves and cavitation, rather than a smooth

simple harmonic
oscillator–type
wave.

Most of the energy carried by high-intensity ultrasound in tissue is converted to thermal energy. In fact, intensities of 10^3 to $10^4\ \text{W/m}^2$ are commonly used for deep-heat treatments called ultrasound diathermy. Frequencies of 0.8 to 1 MHz are typical. In both athletics and physical therapy, ultrasound diathermy is most often applied to injured or overworked muscles to relieve pain and improve flexibility. Skill is needed by the therapist to avoid “bone burns” and other tissue damage caused by overheating and cavitation, sometimes made worse by reflection and focusing of the ultrasound by joint and bone tissue.

In some instances, you may encounter a different decibel scale, called the sound *pressure* level, when ultrasound travels in water or in human and other biological tissues. We shall not use the scale here, but it is notable that numbers for sound pressure levels range 60 to 70 dB higher than you would quote for β, the sound intensity level used in this text. Should you encounter a sound pressure level of 220 decibels, then, it is not an astronomically high intensity, but equivalent to about 155 dB—high enough to destroy tissue, but not as unreasonably high as it might seem at first.

Ultrasound in Medical Diagnostics

When used for imaging, ultrasonic waves are emitted from a transducer, a crystal exhibiting the piezoelectric effect (the expansion and contraction of a substance when a voltage is applied across it, causing a vibration of the crystal). These high-frequency vibrations are transmitted into any tissue in contact with the transducer. Similarly, if a pressure is applied to the crystal (in the form of a wave reflected off tissue layers), a voltage is produced which can be recorded. The crystal therefore acts as both a transmitter and a receiver of sound. Ultrasound is also partially absorbed by tissue on its path, both on its journey away from the transducer and on its return journey. From the time between when the original signal is sent and when the reflections from various boundaries between media are received, (as well as a measure of the intensity loss of the signal), the nature and position of each boundary between tissues and organs may be deduced.

Reflections at boundaries between two different media occur because of differences in a characteristic known as the **acoustic impedance** Z of each substance. Impedance is defined as

Equation:

$$Z = \rho v,$$

where ρ is the density of the medium (in kg/m^3) and v is the speed of sound through the medium (in m/s). The units for Z are therefore $\text{kg/(m}^2 \cdot \text{s)}$.

[link] shows the density and speed of sound through various media (including various soft tissues) and the associated acoustic impedances. Note that the acoustic impedances for soft tissue do not vary much but that there is a big difference between the acoustic impedance of soft tissue and air and also between soft tissue and bone.

Medium	**Density (kg/m^3)**	**Speed of Ultrasound (m/s)**	**Acoustic Impedance** $\left(\text{kg}/\left(\text{m}^2 \cdot \text{s}\right)\right)$
Air	1.3	330	429
Water	1000	1500	1.5×10^6
Blood	1060	1570	1.66×10^6
Fat	925	1450	1.34×10^6
Muscle (average)	1075	1590	1.70×10^6
Bone (varies)	1400–1900	4080	5.7×10^6 to 7.8×10^6
Barium titanate (transducer material)	5600	5500	30.8×10^6

The Ultrasound Properties of Various Media, Including Soft Tissue Found in the Body

At the boundary between media of different acoustic impedances, some of the wave energy is reflected and some is transmitted. The greater the *difference* in acoustic impedance between the two media, the greater the reflection and the smaller the transmission.

The **intensity reflection coefficient** a is defined as the ratio of the intensity of the reflected wave relative to the incident (transmitted) wave. This statement can be written mathematically as

Equation:

$$a=\frac{(Z_2-Z_1)^2}{(Z_1+Z_2)^2},$$

where Z_1 and Z_2 are the acoustic impedances of the two media making up the boundary. A reflection coefficient of zero (corresponding to total transmission and no reflection) occurs when the acoustic impedances of the two media are the same. An impedance “match” (no reflection) provides an efficient coupling of sound energy from one medium to another. The image formed in an ultrasound is made by tracking reflections (as shown in [link]) and mapping the intensity of the reflected sound waves in a two-dimensional plane.

Example:

Calculate Acoustic Impedance and Intensity Reflection Coefficient: Ultrasound and Fat Tissue

(a) Using the values for density and the speed of ultrasound given in [link], show that the acoustic impedance of fat tissue is indeed $1.34\times10^6\ \text{kg}/(\text{m}^2\cdot\text{s})$.

(b) Calculate the intensity reflection coefficient of ultrasound when going from fat to muscle tissue.

Strategy for (a)

The acoustic impedance can be calculated using $Z=\rho v$ and the values for ρ and v found in [link].

Solution for (a)

(1) Substitute known values from [link] into $Z=\rho v$.

Equation:

$$Z=\rho v=\left(925\ \text{kg/m}^3\right)(1450\ \text{m/s})$$

(2) Calculate to find the acoustic impedance of fat tissue.

Equation:

$$1.34\times10^{6}\ \text{kg/(m}^{2}\cdot\text{s)}$$

This value is the same as the value given for the acoustic impedance of fat tissue.

Strategy for (b)

The intensity reflection coefficient for any boundary between two media is given by $a=\frac{(Z_2-Z_1)^2}{(Z_1+Z_2)^2}$, and the acoustic impedance of muscle is given in [link].

Solution for (b)

Substitute known values into $a=\frac{(Z_2-Z_1)^2}{(Z_1+Z_2)^2}$ to find the intensity reflection coefficient:

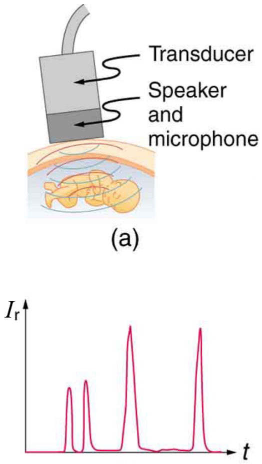
Equation:

$$a=\frac{(Z_2-Z_1)^2}{(Z_1+Z_2)^2}=\frac{\left(1.34\times10^{6}\ \text{kg/(m}^{2}\cdot\text{ s)}-1.70\times10^{6}\ \text{kg/(m}^{2}\cdot\text{ s)}\right)^2}{\left(1.70\times10^{6}\ \text{kg/(m}^{2}\cdot\text{ s)}+1.34\times10^{6}\ \text{kg/(m}^{2}\cdot\text{ s)}\right)^2}=0.014$$

Discussion

This result means that only 1.4% of the incident intensity is reflected, with the remaining being transmitted.

The applications of ultrasound in medical diagnostics have produced untold benefits with no known risks. Diagnostic intensities are too low (about $10^{-2}\ \text{W/m}^{2}$) to cause thermal damage. More significantly, ultrasound has been in use for several decades and detailed follow-up studies do not show evidence of ill effects, quite unlike the case for x-rays.

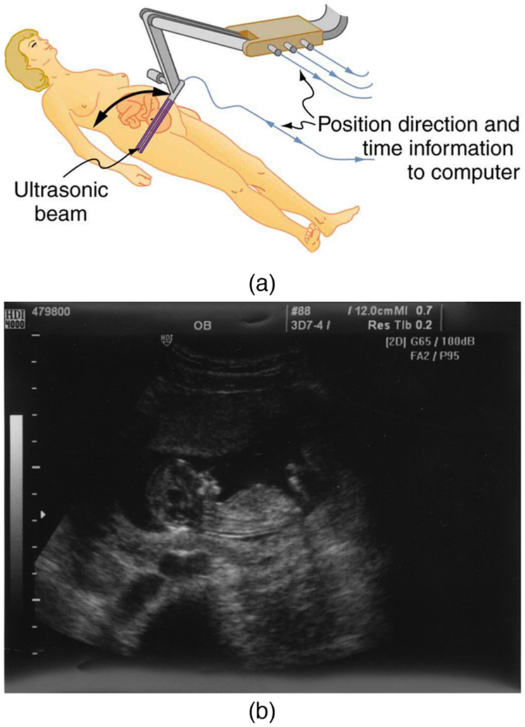

Transducer
Speaker and microphone
I_r
t

(a)

(b)

(a) An ultrasound speaker doubles as a microphone. Brief bleeps are broadcast, and echoes are recorded from various depths. (b) Graph of echo intensity versus time. The time for echoes to return is directly proportional to the distance of the reflector, yielding this information noninvasively.

The most common ultrasound applications produce an image like that shown in [link]. The speaker-microphone broadcasts a directional beam, sweeping the beam across the area of interest. This is accomplished by having multiple ultrasound sources in the probe’s head, which are phased to interfere constructively in a given, adjustable direction. Echoes are measured as a function of position as well as depth. A computer constructs an image that reveals the shape and density of internal structures.


Position direction and time information to computer
Ultrasonic beam
(a)
479800
OB
#88
3D7-4 /
/ 12.0cmMI 0.7
Res Tlb 0.2
[2D] G65 / 100dB
FA2 / P95
(b)

(a) An ultrasonic image is produced by sweeping the ultrasonic beam across the area of interest, in this case the woman’s abdomen. Data are recorded and analyzed in a computer, providing a two-dimensional image. (b) Ultrasound image of 12-week-old fetus. (credit: Margaret W. Carruthers, Flickr)

How much detail can ultrasound reveal? The image in [link] is typical of low-cost systems, but that in [link] shows the remarkable detail possible with more advanced systems, including 3D imaging. Ultrasound today is commonly used in prenatal care. Such imaging can be used to see if the fetus is developing at a normal rate, and help in the determination of serious problems early in the pregnancy. Ultrasound is also in wide use to image the chambers of the heart and the flow of blood within the beating heart, using the Doppler effect (echocardiology).

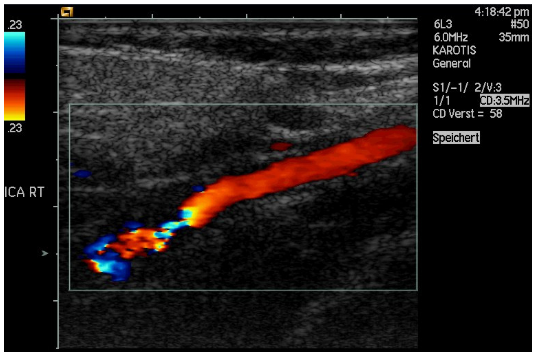
Whenever a wave is used as a probe, it is very difficult to detect details smaller than its wavelength λ. Indeed, current technology cannot do quite this well. Abdominal scans may use a 7-MHz frequency, and the speed of sound in tissue is about 1540 m/s—so the wavelength limit to detail would be $\lambda = \frac{v_{\text{w}}}{f} = \frac{1540\text{ m/s}}{7\times10^{6}\text{ Hz}} = 0.22\text{ mm}$. In practice, 1-mm detail is attainable, which is sufficient for many purposes. Higher-frequency ultrasound would allow greater detail, but it does not penetrate as well as lower frequencies do. The accepted rule of thumb is that you can effectively scan to a depth of about 500λ into tissue. For 7 MHz, this penetration limit is $500 \times 0.22\text{ mm}$, which is 0.11 m. Higher frequencies may be employed in smaller organs, such as the eye, but are not practical for looking deep into the body.

4D Fetal Imaging (408)249-2229
10-02-2007
CU, J
RAB 4-8P/Obstetric
2.8/9.6cm
1.5Hz
3:29:19 PM

A 3D ultrasound image of a fetus. As well as for the detection of any abnormalities, such scans have also been shown to be useful for strengthening the emotional bonding between parents and their unborn child. (credit: Jennie Cu, Wikimedia Commons)

In addition to shape information, ultrasonic scans can produce density information superior to that found in X-rays, because the intensity of a reflected sound is related to changes in density. Sound is most strongly reflected at places where density changes are greatest.

Another major use of ultrasound in medical diagnostics is to detect motion and determine velocity through the Doppler shift of an echo, known as **Doppler-shifted ultrasound**. This technique is used to monitor fetal heartbeat, measure blood velocity, and detect occlusions in blood vessels, for example. (See [link].) The magnitude of the Doppler shift in an echo is directly proportional to the velocity of whatever reflects the sound. Because an echo is involved, there is actually a double shift. The first occurs because the reflector (say a fetal heart) is a moving observer and receives a Doppler-shifted frequency. The reflector then acts as a moving source, producing a second Doppler shift.



This Doppler-shifted ultrasonic image of a partially occluded artery uses color to indicate velocity. The highest velocities are in red, while the lowest are blue. The blood must move faster through the constriction to carry the same flow. (credit: Arning C, Grzyska U, Wikimedia Commons)

A clever technique is used to measure the Doppler shift in an echo. The frequency of the echoed sound is superimposed on the broadcast frequency, producing beats. The beat frequency is $F_{\text{B}} = | f_1 - f_2 |$, and so it is directly proportional to the Doppler shift $(f_1 - f_2)$ and hence, the reflector's velocity. The advantage in this technique is that the Doppler shift is small (because the reflector's velocity is small), so that great accuracy would be needed to measure the shift directly. But measuring the beat frequency is easy, and it is not affected if the broadcast frequency varies somewhat. Furthermore, the beat frequency is in the audible range and can be amplified for audio feedback to the medical observer.

Note:

Uses for Doppler-Shifted Radar

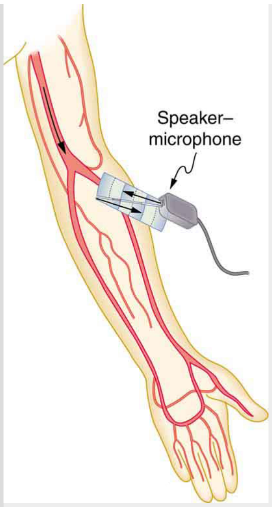
Doppler-shifted radar echoes are used to measure wind velocities in storms as well as aircraft and automobile speeds. The principle is the same as for Doppler-shifted ultrasound. There is evidence that bats and dolphins may also sense the velocity of an object (such as prey) reflecting their ultrasound signals by observing its Doppler shift.

Example:

Calculate Velocity of Blood: Doppler-Shifted Ultrasound

Ultrasound that has a frequency of 2.50 MHz is sent toward blood in an artery that is moving toward the source at 20.0 cm/s, as illustrated in [link]. Use the speed of sound in human tissue as 1540 m/s. (Assume that the frequency of 2.50 MHz is accurate to seven significant figures.)

a. What frequency does the blood receive?
b. What frequency returns to the source?
c. What beat frequency is produced if the source and returning frequencies are mixed?



Ultrasound is partly reflected by blood cells and plasma back toward the speaker-microphone. Because the cells are moving, two Doppler shifts are produced—one for blood as a moving observer, and the other for the reflected sound coming from a moving source. The magnitude of the shift is directly proportional to blood velocity.

Strategy

The first two questions can be answered using $f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} \pm v_{\text{s}}}\right)$ and $f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}} \pm v_{\text{obs}}}{v_{\text{w}}}\right)$ for the Doppler shift. The last question asks for beat frequency, which is the difference between the original and returning frequencies.

Solution for (a)

(1) Identify knowns:

- The blood is a moving observer, and so the frequency it receives is given by **Equation:**

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}} \pm v_{\text{obs}}}{v_{\text{w}}}\right).$$

- v_{b} is the blood velocity (v_{obs} here) and the plus sign is chosen because the motion is toward the source.

(2) Enter the given values into the equation.

Equation:

$$f_{\text{obs}} = (2{,}500{,}000\ \text{Hz})\left(\frac{1540\ \text{m/s} + 0.2\ \text{m/s}}{1540\ \text{m/s}}\right)$$

(3) Calculate to find the frequency: 2,500,325 Hz.

Solution for (b)

(1) Identify knowns:

- The blood acts as a moving source.
- The microphone acts as a stationary observer.
- The frequency leaving the blood is 2,500,325 Hz, but it is shifted upward as given by
  Equation:

$$f_{\text{obs}} = f_{\text{s}}\left(\frac{v_{\text{w}}}{v_{\text{w}} - v_{\text{b}}}\right).$$

  f_{obs} is the frequency received by the speaker-microphone.
- The source velocity is v_{b}.
- The minus sign is used because the motion is toward the observer.

The minus sign is used because the motion is toward the observer.
(2) Enter the given values into the equation:
Equation:

$$f_{\text{obs}} = (2{,}500{,}325\ \text{Hz})\left(\frac{1540\ \text{m/s}}{1540\ \text{m/s} - 0.200\ \text{m/s}}\right)$$

(3) Calculate to find the frequency returning to the source: 2,500,649 Hz.
Solution for (c)
(1) Identify knowns:

- The beat frequency is simply the absolute value of the difference between f_{s} and f_{obs}, as stated in:
  Equation:

  $$f_{\text{B}} = \mid f_{\text{obs}} - f_{\text{s}} \mid .$$

(2) Substitute known values:
Equation:

$$\mid 2{,}500{,}649\ \text{Hz} - 2{,}500{,}000\ \text{Hz} \mid$$

(3) Calculate to find the beat frequency: 649 Hz.
Discussion
The Doppler shifts are quite small compared with the original frequency of 2.50 MHz. It is far easier to measure the beat frequency than it is to measure the echo frequency with an accuracy great enough to see shifts of a few hundred hertz out of a couple of megahertz. Furthermore, variations in the source frequency do not greatly affect the beat frequency, because both f_{s} and f_{obs}would increase or decrease. Those changes subtract out in $f_{\text{B}} = \mid f_{\text{obs}} - f_{\text{s}} \mid$.

Note:
Industrial and Other Applications of Ultrasound
Industrial, retail, and research applications of ultrasound are common. A few are discussed here. Ultrasonic cleaners have many uses. Jewelry, machined parts, and other objects that have odd shapes and crevices are immersed in a cleaning fluid that is agitated with ultrasound typically about 40 kHz in frequency. The intensity is great enough to cause cavitation, which is responsible for most of the cleansing action. Because cavitation-produced shock pressures are large and well transmitted in a fluid,

they reach into small crevices where even a low-surface-tension cleaning fluid might not penetrate.

Sonar is a familiar application of ultrasound. Sonar typically employs ultrasonic frequencies in the range from 30.0 to 100 kHz. Bats, dolphins, submarines, and even some birds use ultrasonic sonar. Echoes are analyzed to give distance and size information both for guidance and finding prey. In most sonar applications, the sound reflects quite well because the objects of interest have significantly different density than the medium in which they travel. When the Doppler shift is observed, velocity information can also be obtained. Submarine sonar can be used to obtain such information, and there is evidence that some bats also sense velocity from their echoes.

Similarly, there are a range of relatively inexpensive devices that measure distance by timing ultrasonic echoes. Many cameras, for example, use such information to focus automatically. Some doors open when their ultrasonic ranging devices detect a nearby object, and certain home security lights turn on when their ultrasonic rangers observe motion. Ultrasonic “measuring tapes” also exist to measure such things as room dimensions. Sinks in public restrooms are sometimes automated with ultrasound devices to turn faucets on and off when people wash their hands. These devices reduce the spread of germs and can conserve water.

Ultrasound is used for nondestructive testing in industry and by the military. Because ultrasound reflects well from any large change in density, it can reveal cracks and voids in solids, such as aircraft wings, that are too small to be seen with x-rays. For similar reasons, ultrasound is also good for measuring the thickness of coatings, particularly where there are several layers involved.

Basic research in solid state physics employs ultrasound. Its attenuation is related to a number of physical characteristics, making it a useful probe. Among these characteristics are structural changes such as those found in liquid crystals, the transition of a material to a superconducting phase, as well as density and other properties.

These examples of the uses of ultrasound are meant to whet the appetites of the curious, as well as to illustrate the underlying physics of ultrasound. There are many more applications, as you can easily discover for yourself.

Exercise:
Check Your Understanding

Problem:

Why is it possible to use ultrasound both to observe a fetus in the womb and also to destroy cancerous tumors in the body?

Solution:

Ultrasound can be used medically at different intensities. Lower intensities do not cause damage and are used for medical imaging. Higher intensities can pulverize and destroy targeted substances in the body, such as tumors.

Section Summary

- The acoustic impedance is defined as:
  Equation:
  $$Z = \rho v,$$
  ρ is the density of a medium through which the sound travels and v is the speed of sound through that medium.
- The intensity reflection coefficient a, a measure of the ratio of the intensity of the wave reflected off a boundary between two media relative to the intensity of the incident wave, is given by
  Equation:
  $$a = \frac{(Z_2 - Z_1)^2}{(Z_1 + Z_2)^2}.$$
- The intensity reflection coefficient is a unitless quantity.

Conceptual Questions

Exercise:

Problem:

If audible sound follows a rule of thumb similar to that for ultrasound, in terms of its absorption, would you expect the high or low frequencies from your neighbor’s stereo to penetrate into your house? How does this expectation compare with your experience?

Exercise:

Problem:

Elephants and whales are known to use infrasound to communicate over very large distances. What are the advantages of infrasound for long distance communication?

Exercise:

Problem:

It is more difficult to obtain a high-resolution ultrasound image in the abdominal region of someone who is overweight than for someone who has a slight build. Explain why this statement is accurate.

Exercise:

Problem:

Suppose you read that 210-dB ultrasound is being used to pulverize cancerous tumors. You calculate the intensity in watts per centimeter squared and find it is unreasonably high ($10^5\ \mathrm{W/cm^2}$). What is a possible explanation?

Problems & Exercises

Unless otherwise indicated, for problems in this section, assume that the speed of sound through human tissues is 1540 m/s.

Exercise:

Problem:

What is the sound intensity level in decibels of ultrasound of intensity $10^5\ \mathrm{W/m^2}$, used to pulverize tissue during surgery?

Solution:

170 dB

Exercise:

Problem:

Is 155-dB ultrasound in the range of intensities used for deep heating? Calculate the intensity of this ultrasound and compare this intensity with values quoted in the text.

Exercise:

Problem:

Find the sound intensity level in decibels of $2.00 \times 10^{-2}\ \mathrm{W/m^2}$ ultrasound used in medical diagnostics.

Solution:

103 dB

Exercise:

Problem:

The time delay between transmission and the arrival of the reflected wave of a signal using ultrasound traveling through a piece of fat tissue was 0.13 ms. At what depth did this reflection occur?

Exercise:

Problem:

In the clinical use of ultrasound, transducers are always coupled to the skin by a thin layer of gel or oil, replacing the air that would otherwise exist between the transducer and the skin. (a) Using the values of acoustic impedance given in [link] calculate the intensity reflection coefficient between transducer material and air. (b) Calculate the intensity reflection coefficient between transducer material and gel (assuming for this problem that its acoustic impedance is identical to that of water). (c) Based on the results of your calculations, explain why the gel is used.

Solution:

(a) 1.00

(b) 0.823

(c) Gel is used to facilitate the transmission of the ultrasound between the transducer and the patient’s body.

Exercise:

Problem:

(a) Calculate the minimum frequency of ultrasound that will allow you to see details as small as 0.250 mm in human tissue. (b) What is the effective depth to which this sound is effective as a diagnostic probe?

Exercise:

Problem:

(a) Find the size of the smallest detail observable in human tissue with 20.0-MHz ultrasound. (b) Is its effective penetration depth great enough to examine the entire eye (about 3.00 cm is needed)? (c) What is the wavelength of such ultrasound in 0°C air?

Solution:

(a) $77.0\ \mu\text{m}$

(b) Effective penetration depth = 3.85 cm, which is enough to examine the eye.

(c) $16.6\ \mu\text{m}$

Exercise:

Problem:

(a) Echo times are measured by diagnostic ultrasound scanners to determine distances to reflecting surfaces in a patient. What is the difference in echo times for tissues that are 3.50 and 3.60 cm beneath the surface? (This difference is the minimum resolving time for the scanner to see details as small as 0.100 cm, or 1.00 mm. Discrimination of smaller time differences is needed to see smaller details.) (b) Discuss whether the period T of this ultrasound must be smaller than the minimum time resolution. If so, what is the minimum frequency of the ultrasound and is that out of the normal range for diagnostic ultrasound?

Exercise:

Problem:

(a) How far apart are two layers of tissue that produce echoes having round-trip times (used to measure distances) that differ by $0.750\ \mu\text{s}$? (b) What minimum frequency must the ultrasound have to see detail this small?

Solution:

(a) $5.78 \times 10^{-4}\ \text{m}$

(b) $2.67 \times 10^{6}\ \text{Hz}$

Exercise:

Problem:

(a) A bat uses ultrasound to find its way among trees. If this bat can detect echoes 1.00 ms apart, what minimum distance between objects can it detect? (b) Could this distance explain the difficulty that bats have finding an open door when they accidentally get into a house?

Exercise:

Problem:

A dolphin is able to tell in the dark that the ultrasound echoes received from two sharks come from two different objects only if the sharks are separated by 3.50 m, one being that much farther away than the other. (a) If the ultrasound has a frequency of 100 kHz, show this ability is not limited by its wavelength. (b) If this ability is due to the dolphin's ability to detect the arrival times of echoes, what is the minimum time difference the dolphin can perceive?

Solution:

(a) $v_{\text{w}} = 1540\text{ m/s} = f\lambda \Rightarrow \lambda = \frac{1540\text{ m/s}}{100\times 10^3\text{ Hz}} = 0.0154\text{ m} < 3.50\text{ m}$. Because the wavelength is much shorter than the distance in question, the wavelength is not the limiting factor.

(b) 4.55 ms

Exercise:

Problem:

A diagnostic ultrasound echo is reflected from moving blood and returns with a frequency 500 Hz higher than its original 2.00 MHz. What is the velocity of the blood? (Assume that the frequency of 2.00 MHz is accurate to seven significant figures and 500 Hz is accurate to three significant figures.)

Exercise:

Problem:

Ultrasound reflected from an oncoming bloodstream that is moving at 30.0 cm/s is mixed with the original frequency of 2.50 MHz to produce beats. What is the beat frequency? (Assume that the frequency of 2.50 MHz is accurate to seven significant figures.)

Solution:

974 Hz

(Note: extra digits were retained in order to show the difference.)

Glossary

acoustic impedance	property of medium that makes the propagation of sound waves more difficult
intensity reflection coefficient	a measure of the ratio of the intensity of the wave reflected off a boundary between two media relative to the intensity of the incident wave
Doppler-shifted ultrasound	a medical technique to detect motion and determine velocity through the Doppler shift of an echo

Atomic Masses

Atomic Number, *Z*	**Name**	**Atomic Mass Number, *A***	**Symbol**	**Atomic Mass (u)**	**Percent Abundance or Decay Mode**	**Half-life, $t_{1/2}$**
0	neutron	1	n	1.008 665	β^-	10.37 min
1	Hydrogen	1	^1H	1.007 825	99.985%	
	Deuterium	2	^2H or D	2.014 102	0.015%	
	Tritium	3	^3H or T	3.016 050	β^-	12.33 y
2	Helium	3	^3He	3.016 030	$1.38 \times 10^{-4}\%$	
		4	^4He	4.002 603	$\approx 100\%$	
3	Lithium	6	^6Li	6.015 121	7.5%	
		7	^7Li	7.016 003	92.5%	
4	Beryllium	7	^7Be	7.016 928	EC	53.29 d
		9	^9Be	9.012 182	100%	
5	Boron	10	^{10}B	10.012 937	19.9%	
		11	^{11}B	11.009 305	80.1%	
6	Carbon	11	^{11}C	11.011 432	EC, β^+	
		12	^{12}C	12.000 000	98.90%	
		13	^{13}C	13.003 355	1.10%	

Atomic Number, *Z*	**Name**	**Atomic Mass Number, *A***	**Symbol**	**Atomic Mass (u)**	**Percent Abundance or Decay Mode**	**Half-life, $t_{1/2}$**
		14	^{14}C	14.003 241	β^-	5730 y
7	Nitrogen	13	^{13}N	13.005 738	β^+	9.96 min
		14	^{14}N	14.003 074	99.63%	
		15	^{15}N	15.000 108	0.37%	
8	Oxygen	15	^{15}O	15.003 065	EC, β^+	122 s
		16	^{16}O	15.994 915	99.76%	
		18	^{18}O	17.999 160	0.200%	
9	Fluorine	18	^{18}F	18.000 937	EC, β^+	1.83 h
		19	^{19}F	18.998 403	100%	
10	Neon	20	^{20}Ne	19.992 435	90.51%	
		22	^{22}Ne	21.991 383	9.22%	
11	Sodium	22	^{22}Na	21.994 434	β^+	2.602 y
		23	^{23}Na	22.989 767	100%	
		24	^{24}Na	23.990 961	β^-	14.96 h
12	Magnesium	24	^{24}Mg	23.985 042	78.99%	
13	Aluminum	27	^{27}Al	26.981 539	100%	
14	Silicon	28	^{28}Si	27.976 927	92.23%	2.62h

Atomic Number, *Z*	**Name**	**Atomic Mass Number, *A***	**Symbol**	**Atomic Mass (u)**	**Percent Abundance or Decay Mode**	**Half-life, $t_{1/2}$**
		31	^{31}Si	30.975 362	β^-	
15	Phosphorus	31	^{31}P	30.973 762	100%	
		32	^{32}P	31.973 907	β^-	14.28 d
16	Sulfur	32	^{32}S	31.972 070	95.02%	
		35	^{35}S	34.969 031	β^-	87.4 d
17	Chlorine	35	^{35}Cl	34.968 852	75.77%	
		37	^{37}Cl	36.965 903	24.23%	
18	Argon	40	^{40}Ar	39.962 384	99.60%	
19	Potassium	39	^{39}K	38.963 707	93.26%	
		40	^{40}K	39.963 999	0.0117%, EC, β^-	$1.28 \times 10^9\text{y}$
20	Calcium	40	^{40}Ca	39.962 591	96.94%	
21	Scandium	45	^{45}Sc	44.955 910	100%	
22	Titanium	48	^{48}Ti	47.947 947	73.8%	
23	Vanadium	51	^{51}V	50.943 962	99.75%	
24	Chromium	52	^{52}Cr	51.940 509	83.79%	
25	Manganese	55	^{55}Mn	54.938 047	100%	
26	Iron	56	^{56}Fe	55.934 939	91.72%	

Atomic Number, *Z*	Name	Atomic Mass Number, *A*	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
27	Cobalt	59	^{59}Co	58.933 198	100%	
		60	^{60}Co	59.933 819	β^-	5.271 y
28	Nickel	58	^{58}Ni	57.935 346	68.27%	
		60	^{60}Ni	59.930 788	26.10%	
29	Copper	63	^{63}Cu	62.939 598	69.17%	
		65	^{65}Cu	64.927 793	30.83%	
30	Zinc	64	^{64}Zn	63.929 145	48.6%	
		66	^{66}Zn	65.926 034	27.9%	
31	Gallium	69	^{69}Ga	68.925 580	60.1%	
32	Germanium	72	^{72}Ge	71.922 079	27.4%	
		74	^{74}Ge	73.921 177	36.5%	
33	Arsenic	75	^{75}As	74.921 594	100%	
34	Selenium	80	^{80}Se	79.916 520	49.7%	
35	Bromine	79	^{79}Br	78.918 336	50.69%	
36	Krypton	84	^{84}Kr	83.911 507	57.0%	
37	Rubidium	85	^{85}Rb	84.911 794	72.17%	
38	Strontium	86	^{86}Sr	85.909 267	9.86%	

Atomic Number, *Z*	**Name**	**Atomic Mass Number, *A***	**Symbol**	**Atomic Mass (u)**	**Percent Abundance or Decay Mode**	**Half-life, $t_{1/2}$**
		88	^{88}Sr	87.905 619	82.58%	
		90	^{90}Sr	89.907 738	β^-	28.8 y
39	Yttrium	89	^{89}Y	88.905 849	100%	
		90	^{90}Y	89.907 152	β^-	64.1 h
40	Zirconium	90	^{90}Zr	89.904 703	51.45%	
41	Niobium	93	^{93}Nb	92.906 377	100%	
42	Molybdenum	98	^{98}Mo	97.905 406	24.13%	
43	Technetium	98	^{98}Tc	97.907 215	β^-	$4.2 \times 10^6\text{y}$
44	Ruthenium	102	^{102}Ru	101.904 348	31.6%	
45	Rhodium	103	^{103}Rh	102.905 500	100%	
46	Palladium	106	^{106}Pd	105.903 478	27.33%	
47	Silver	107	^{107}Ag	106.905 092	51.84%	
		109	^{109}Ag	108.904 757	48.16%	
48	Cadmium	114	^{114}Cd	113.903 357	28.73%	
49	Indium	115	^{115}In	114.903 880	95.7%, β^-	$4.4 \times 10^{14}\text{y}$
50	Tin	120	^{120}Sn	119.902 200	32.59%	
51	Antimony	121	^{121}Sb	120.903 821	57.3%	

Atomic Number, *Z*	**Name**	**Atomic Mass Number, *A***	**Symbol**	**Atomic Mass (u)**	**Percent Abundance or Decay Mode**	**Half-life, $t_{1/2}$**
52	Tellurium	130	^{130}Te	129.906 229	33.8%, β^-	$2.5 \times 10^{21}\text{y}$
53	Iodine	127	^{127}I	126.904 473	100%	
		131	^{131}I	130.906 114	β^-	8.040 d
54	Xenon	132	^{132}Xe	131.904 144	26.9%	
		136	^{136}Xe	135.907 214	8.9%	
55	Cesium	133	^{133}Cs	132.905 429	100%	
		134	^{134}Cs	133.906 696	EC, β^-	2.06 y
56	Barium	137	^{137}Ba	136.905 812	11.23%	
		138	^{138}Ba	137.905 232	71.70%	
57	Lanthanum	139	^{139}La	138.906 346	99.91%	
58	Cerium	140	^{140}Ce	139.905 433	88.48%	
59	Praseodymium	141	^{141}Pr	140.907 647	100%	
60	Neodymium	142	^{142}Nd	141.907 719	27.13%	
61	Promethium	145	^{145}Pm	144.912 743	EC, α	17.7 y
62	Samarium	152	^{152}Sm	151.919 729	26.7%	
63	Europium	153	^{153}Eu	152.921 225	52.2%	
64	Gadolinium	158	^{158}Gd	157.924 099	24.84%	

Atomic Number, *Z*	Name	Atomic Mass Number, *A*	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
65	Terbium	159	^{159}Tb	158.925 342	100%	
66	Dysprosium	164	^{164}Dy	163.929 171	28.2%	
67	Holmium	165	^{165}Ho	164.930 319	100%	
68	Erbium	166	^{166}Er	165.930 290	33.6%	
69	Thulium	169	^{169}Tm	168.934 212	100%	
70	Ytterbium	174	^{174}Yb	173.938 859	31.8%	
71	Lutecium	175	^{175}Lu	174.940 770	97.41%	
72	Hafnium	180	^{180}Hf	179.946 545	35.10%	
73	Tantalum	181	^{181}Ta	180.947 992	99.98%	
74	Tungsten	184	^{184}W	183.950 928	30.67%	
75	Rhenium	187	^{187}Re	186.955 744	62.6%, β^-	$4.6 \times 10^{10}\text{y}$
76	Osmium	191	^{191}Os	190.960 920	β^-	15.4 d
		192	^{192}Os	191.961 467	41.0%	
77	Iridium	191	^{191}Ir	190.960 584	37.3%	
		193	^{193}Ir	192.962 917	62.7%	
78	Platinum	195	^{195}Pt	194.964 766	33.8%	
79	Gold	197	^{197}Au	196.966 543	100%	

Atomic Number, *Z*	**Name**	**Atomic Mass Number, *A***	**Symbol**	**Atomic Mass (u)**	**Percent Abundance or Decay Mode**	**Half-life, $t_{1/2}$**
		198	^{198}Au	197.968 217	β^-	2.696 d
80	Mercury	199	^{199}Hg	198.968 253	16.87%	
		202	^{202}Hg	201.970 617	29.86%	
81	Thallium	205	^{205}Tl	204.974 401	70.48%	
82	Lead	206	^{206}Pb	205.974 440	24.1%	
		207	^{207}Pb	206.975 872	22.1%	
		208	^{208}Pb	207.976 627	52.4%	
		210	^{210}Pb	209.984 163	α, β^-	22.3 y
		211	^{211}Pb	210.988 735	β^-	36.1 min
		212	^{212}Pb	211.991 871	β^-	10.64 h
83	Bismuth	209	^{209}Bi	208.980 374	100%	
		211	^{211}Bi	210.987 255	α, β^-	2.14 min
84	Polonium	210	^{210}Po	209.982 848	α	138.38 d
85	Astatine	218	^{218}At	218.008 684	α, β^-	1.6 s
86	Radon	222	^{222}Rn	222.017 570	α	3.82 d
87	Francium	223	^{223}Fr	223.019 733	α, β^-	21.8 min
88	Radium	226	^{226}Ra	226.025 402	α	1.60×10^3y

Atomic Number, *Z*	**Name**	**Atomic Mass Number, *A***	**Symbol**	**Atomic Mass (u)**	**Percent Abundance or Decay Mode**	**Half-life, $t_{1/2}$**
89	Actinium	227	^{227}Ac	227.027 750	α, β^-	21.8 y
90	Thorium	228	^{228}Th	228.028 715	α	1.91 y
		232	^{232}Th	232.038 054	100%, α	$1.41 \times 10^{10}\text{y}$
91	Protactinium	231	^{231}Pa	231.035 880	α	$3.28 \times 10^{4}\text{y}$
92	Uranium	233	^{233}U	233.039 628	α	$1.59 \times 10^{3}\text{y}$
		235	^{235}U	235.043 924	0.720%, α	$7.04 \times 10^{8}\text{y}$
		236	^{236}U	236.045 562	α	$2.34 \times 10^{7}\text{y}$
		238	^{238}U	238.050 784	99.2745%, α	$4.47 \times 10^{9}\text{y}$
		239	^{239}U	239.054 289	β^-	23.5 min
93	Neptunium	239	^{239}Np	239.052 933	β^-	2.355 d
94	Plutonium	239	^{239}Pu	239.052 157	α	$2.41 \times 10^{4}\text{y}$
95	Americium	243	^{243}Am	243.061 375	α, fission	$7.37 \times 10^{3}\text{y}$
96	Curium	245	^{245}Cm	245.065 483	α	$8.50 \times 10^{3}\text{y}$
97	Berkelium	247	^{247}Bk	247.070 300	α	$1.38 \times 10^{3}\text{y}$
98	Californium	249	^{249}Cf	249.074 844	α	351 y
99	Einsteinium	254	^{254}Es	254.088 019	α, β^-	276 d
100	Fermium	253	^{253}Fm	253.085 173	EC, α	3.00 d

Atomic Number, *Z*	Name	Atomic Mass Number, *A*	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
101	Mendelevium	255	^{255}Md	255.091 081	EC, α	27 min
102	Nobelium	255	^{255}No	255.093 260	EC, α	3.1 min
103	Lawrencium	257	^{257}Lr	257.099 480	EC, α	0.646 s
104	Rutherfordium	261	^{261}Rf	261.108 690	α	1.08 min
105	Dubnium	262	^{262}Db	262.113 760	α, fission	34 s
106	Seaborgium	263	^{263}Sg	263.11 86	α, fission	0.8 s
107	Bohrium	262	^{262}Bh	262.123 1	α	0.102 s
108	Hassium	264	^{264}Hs	264.128 5	α	0.08 ms
109	Meitnerium	266	^{266}Mt	266.137 8	α	3.4 ms

Atomic Masses

Useful Information

This appendix is broken into several tables.

- [link], Important Constants
- [link], Submicroscopic Masses
- [link], Solar System Data
- [link], Metric Prefixes for Powers of Ten and Their Symbols
- [link], The Greek Alphabet
- [link], SI units
- [link], Selected British Units
- [link], Other Units
- [link], Useful Formulae

Symbol	Meaning	Best Value	Approximate Value
c	Speed of light in vacuum	$2.99792458 \times 10^{8}\,\text{m/s}$	$3.00 \times 10^{8}\,\text{m/s}$
G	Gravitational constant	$6.67408(31) \times 10^{-11}\,\text{N}\cdot\text{m}^2/\text{kg}^2$	$6.67 \times 10^{-11}\,\text{N}\cdot\text{m}^2/\text{kg}^2$
N_A	Avogadro’s number	$6.02214129(27) \times 10^{23}$	6.02×10^{23}
k	Boltzmann’s constant	$1.3806488(13) \times 10^{-23}\,\text{J/K}$	$1.38 \times 10^{-23}\,\text{J/K}$
R	Gas constant	$8.3144621(75)\,\text{J/mol}\cdot\text{K}$	$8.31\,\text{J/mol}\cdot\text{K} = 1.99\,\text{cal/mol}\cdot\text{K} =$
σ	Stefan-Boltzmann constant	$5.670373(21) \times 10^{-8}\,\text{W/m}^2\cdot\text{K}$	$5.67 \times 10^{-8}\,\text{W/m}^2\cdot\text{K}$
k	Coulomb force constant	$8.987551788\ldots \times 10^{9}\,\text{N}\cdot\text{m}^2/\text{C}^2$	$8.99 \times 10^{9}\,\text{N}\cdot\text{m}^2/\text{C}^2$
q_e	Charge on electron	$-1.602176565(35) \times 10^{-19}\,\text{C}$	$-1.60 \times 10^{-19}\,\text{C}$
ε_0	Permittivity of free space	$8.854187817\ldots \times 10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2$	$8.85 \times 10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2$
μ_0	Permeability of free space	$4\pi \times 10^{-7}\,\text{T}\cdot\text{m/A}$	$1.26 \times 10^{-6}\,\text{T}\cdot\text{m/A}$
h	Planck’s constant	$6.62606957(29) \times 10^{-34}\,\text{J}\cdot\text{s}$	$6.63 \times 10^{-34}\,\text{J}\cdot\text{s}$

Important Constants[footnote]

Stated values are according to the National Institute of Standards and Technology Reference on Constants, Units, an
www.physics.nist.gov/cuu (accessed May 18, 2012). Values in parentheses are the uncertainties in the last digits. Nu
are exact as defined.

Symbol	Meaning	Best Value	Approximate Value
m_e	Electron mass	$9.10938291(40) \times 10^{-31}\text{kg}$	$9.11 \times 10^{-31}\text{kg}$
m_p	Proton mass	$1.672621777(74) \times 10^{-27}\text{kg}$	$1.6726 \times 10^{-27}\text{kg}$
m_n	Neutron mass	$1.674927351(74) \times 10^{-27}\text{kg}$	$1.6749 \times 10^{-27}\text{kg}$
u	Atomic mass unit	$1.660538921(73) \times 10^{-27}\text{kg}$	$1.6605 \times 10^{-27}\text{kg}$

Submicroscopic Masses[footnote]
Stated values are according to the National Institute of Standards and Technology Reference on Constants, Units, and Uncertainty, www.physics.nist.gov/cuu (accessed May 18, 2012). Values in parentheses are the uncertainties in the last digits. Numbers without uncertainties are exact as defined.

Sun	mass	$1.99 \times 10^{30}\text{kg}$
	average radius	$6.96 \times 10^{8}\text{m}$
	Earth-sun distance (average)	$1.496 \times 10^{11}\text{m}$
Earth	mass	$5.9736 \times 10^{24}\text{kg}$
	average radius	$6.376 \times 10^{6}\text{m}$
	orbital period	$3.16 \times 10^{7}\text{s}$

Moon	mass	$7.35 \times 10^{22}\text{kg}$
	average radius	$1.74 \times 10^{6}\text{m}$
	orbital period (average)	$2.36 \times 10^{6}\text{s}$
	Earth-moon distance (average)	$3.84 \times 10^{8}\text{m}$

Solar System Data

Prefix	Symbol	Value	Prefix	Symbol	Value
tera	T	10^{12}	deci	d	10^{-1}
giga	G	10^{9}	centi	c	10^{-2}
mega	M	10^{6}	milli	m	10^{-3}
kilo	k	10^{3}	micro	μ	10^{-6}
hecto	h	10^{2}	nano	n	10^{-9}
deka	da	10^{1}	pico	p	10^{-12}
—	—	$10^{0}(=1)$	femto	f	10^{-15}

Metric Prefixes for Powers of Ten and Their Symbols

Alpha	A	α	Eta	H	η	Nu	N	ν	Tau	T	τ
Beta	B	β	Theta	Θ	θ	Xi	Ξ	ξ	Upsilon	Υ	υ
Gamma	Γ	γ	Iota	I	ι	Omicron	O	o	Phi	Φ	ϕ
Delta	Δ	δ	Kappa	K	κ	Pi	Π	π	Chi	X	χ

Epsilon	E	ε	Lambda	Λ	λ	Rho	P	ρ	Psi	Ψ	ψ
Zeta	Z	ζ	Mu	M	μ	Sigma	Σ	σ	Omega	Ω	ω

The Greek Alphabet

	Entity	**Abbreviation**	**Name**
Fundamental units	Length	m	meter
	Mass	kg	kilogram
	Time	s	second
	Current	A	ampere
Supplementary unit	Angle	rad	radian
Derived units	Force	$\mathrm{N} = \mathrm{kg} \cdot \mathrm{m}/\mathrm{s}^2$	newton
	Energy	$\mathrm{J} = \mathrm{kg} \cdot \mathrm{m}^2/\mathrm{s}^2$	joule
	Power	$\mathrm{W} = \mathrm{J}/\mathrm{s}$	watt
	Pressure	$\mathrm{Pa} = \mathrm{N}/\mathrm{m}^2$	pascal
	Frequency	$\mathrm{Hz} = 1/\mathrm{s}$	hertz
	Electronic potential	$\mathrm{V} = \mathrm{J}/\mathrm{C}$	volt
	Capacitance	$\mathrm{F} = \mathrm{C}/\mathrm{V}$	farad
	Charge	$\mathrm{C} = \mathrm{s} \cdot \mathrm{A}$	coulomb
	Resistance	$\Omega = \mathrm{V}/\mathrm{A}$	ohm

	Entity	Abbreviation	Name
	Magnetic field	$\mathrm{T = N/(A \cdot m)}$	tesla
	Nuclear decay rate	$\mathrm{Bq = 1/s}$	becquerel

SI Units

Length	$1\text{ inch (in.)} = 2.54\,\mathrm{cm}\text{ (exactly)}$
	$1\text{ foot (ft)} = 0.3048\,\mathrm{m}$
	$1\text{ mile (mi)} = 1.609\,\mathrm{km}$
Force	$1\text{ pound (lb)} = 4.448\,\mathrm{N}$
Energy	$1\text{ British thermal unit (Btu)} = 1.055 \times 10^3\,\mathrm{J}$
Power	$1\text{ horsepower (hp)} = 746\,\mathrm{W}$
Pressure	$1\,\mathrm{lb/in^2} = 6.895 \times 10^3\,\mathrm{Pa}$

Selected British Units

Length	$1\text{ light year (ly)} = 9.46 \times 10^{15}\,\mathrm{m}$
	$1\text{ astronomical unit (au)} = 1.50 \times 10^{11}\,\mathrm{m}$
	$1\text{ nautical mile} = 1.852\,\mathrm{km}$
	$1\text{ angstrom}(\text{Å}) \ = 10^{-10}\,\mathrm{m}$
Area	$1\text{ acre (ac)} = 4.05 \times 10^3\,\mathrm{m^2}$
	$1\text{ square foot }(\mathrm{ft^2}) = 9.29 \times 10^{-2}\,\mathrm{m^2}$
	$1\text{ barn }(b) = 10^{-28}\,\mathrm{m^2}$
Volume	$1\text{ liter }(L) = 10^{-3}\,\mathrm{m^3}$

	$1\text{ U.S. gallon (gal)} = 3.785 \times 10^{-3}\text{ m}^3$
Mass	$1\text{ solar mass } = 1.99 \times 10^{30}\text{ kg}$
	$1\text{ metric ton} = 10^3\text{ kg}$
	$1\text{ atomic mass unit }(u) = 1.6605 \times 10^{-27}\text{ kg}$
Time	$1\text{ year }(y) = 3.16 \times 10^7\text{ s}$
	$1\text{ day }(d) = 86{,}400\text{ s}$
Speed	$1\text{ mile per hour (mph)} = 1.609\text{ km/h}$
	$1\text{ nautical mile per hour (naut)} = 1.852\text{ km/h}$
Angle	$1\text{ degree }(^\circ) = 1.745 \times 10^{-2}\text{ rad}$
	$1\text{ minute of arc }(') = 1/60\text{ degree}$
	$1\text{ second of arc }('') = 1/60\text{ minute of arc}$
	$1\text{ grad} = 1.571 \times 10^{-2}\text{ rad}$
Energy	$1\text{ kiloton TNT (kT)} = 4.2 \times 10^{12}\text{ J}$
	$1\text{ kilowatt hour }(\text{kW} \cdot h) = 3.60 \times 10^6\text{ J}$
	$1\text{ food calorie (kcal)} = 4186\text{ J}$
	$1\text{ calorie (cal)} = 4.186\text{ J}$
	$1\text{ electron volt (eV)} = 1.60 \times 10^{-19}\text{ J}$
Pressure	$1\text{ atmosphere (atm)} = 1.013 \times 10^5\text{ Pa}$
	$1\text{ millimeter of mercury (mm Hg)} = 133.3\text{ Pa}$
	$1\text{ torricelli (torr)} = 1\text{ mm Hg} = 133.3\text{ Pa}$
Nuclear decay rate	$1\text{ curie (Ci)} = 3.70 \times 10^{10}\text{ Bq}$

Other Units

Circumference of a circle with radius r or diameter d	$C = 2\pi r = \pi d$
Area of a circle with radius r or diameter d	$A = \pi r^2 = \pi d^2/4$
Area of a sphere with radius r	$A = 4\pi r^2$

Volume of a sphere with radius r	$V=(4/3)\left(\pi r^3\right)$

Useful Formulae

Glossary of Key Symbols and Notation

In this glossary, key symbols and notation are briefly defined.

Symbol	Definition
any symbol	average (indicated by a bar over a symbol—e.g., v is average velocity)
$^\circ\text{C}$	Celsius degree
$^\circ\text{F}$	Fahrenheit degree
$//$	parallel
$\perp$	perpendicular
$\propto$	proportional to
$\pm$	plus or minus

Symbol	**Definition**
$_0$	zero as a subscript denotes an initial value
α	alpha rays
α	angular acceleration
α	temperature coefficient(s) of resistivity
β	beta rays
β	sound level
β	volume coefficient of expansion
β^-	electron emitted in nuclear beta decay
β^+	positron decay
γ	gamma rays

Symbol	Definition
γ	surface tension
$\gamma = 1/\sqrt{1 - v^2/c^2}$	a constant used in relativity
Δ	change in whatever quantity follows
δ	uncertainty in whatever quantity follows
$\varDelta E$	change in energy between the initial and final orbits of an electron in an atom
$\varDelta E$	uncertainty in energy
$\varDelta m$	difference in mass between initial and final products
$\varDelta N$	number of decays that occur
$\varDelta p$	change in momentum

Symbol	**Definition**
Δp	uncertainty in momentum
$\Delta \mathrm{PE}_{\mathrm{g}}$	change in gravitational potential energy
$\Delta \theta$	rotation angle
Δs	distance traveled along a circular path
Δt	uncertainty in time
Δt_0	proper time as measured by an observer at rest relative to the process
ΔV	potential difference
Δx	uncertainty in position
ε_0	permittivity of free space
η	viscosity

Symbol	Definition
θ	angle between the force vector and the displacement vector
θ	angle between two lines
θ	contact angle
θ	direction of the resultant
θ_b	Brewster's angle
θ_c	critical angle
κ	dielectric constant
λ	decay constant of a nuclide
λ	wavelength
λ_n	wavelength in a medium

Symbol	**Definition**
μ_0	permeability of free space
μ_k	coefficient of kinetic friction
μ_s	coefficient of static friction
v_e	electron neutrino
π^+	positive pion
π^-	negative pion
π^0	neutral pion
ρ	density
ρ_c	critical density, the density needed to just halt universal expansion
ρ_fl	fluid density

Symbol	Definition
ρ_{obj}	average density of an object
ρ/ρ_{w}	specific gravity
τ	characteristic time constant for a resistance and inductance (RL) or resistance and capacitance (RC) circuit
τ	characteristic time for a resistor and capacitor (RC) circuit
τ	torque
Υ	upsilon meson
Φ	magnetic flux
ϕ	phase angle
Ω	ohm (unit)
ω	angular velocity

Symbol	Definition
A	ampere (current unit)
A	area
A	cross-sectional area
A	total number of nucleons
a	acceleration
a_{B}	Bohr radius
a_{c}	centripetal acceleration
a_{t}	tangential acceleration
AC	alternating current
AM	amplitude modulation

Symbol	Definition
atm	atmosphere
B	baryon number
B	blue quark color
B	antiblue (yellow) antiquark color
b	quark flavor bottom or beauty
B	bulk modulus
B	magnetic field strength
$\mathrm{B_{int}}$	electron’s intrinsic magnetic field
$\mathrm{B_{orb}}$	orbital magnetic field
BE	binding energy of a nucleus—it is the energy required to completely disassemble it into separate protons and neutrons

Symbol	**Definition**
BE/A	binding energy per nucleon
Bq	becquerel—one decay per second
C	capacitance (amount of charge stored per volt)
C	coulomb (a fundamental SI unit of charge)
C_{p}	total capacitance in parallel
C_{s}	total capacitance in series
CG	center of gravity
CM	center of mass
c	quark flavor charm
c	specific heat

Symbol	Definition
c	speed of light
Cal	kilocalorie
cal	calorie
COP_{hp}	heat pump’s coefficient of performance
COP_{ref}	coefficient of performance for refrigerators and air conditioners
$\cos\theta$	cosine
$\cot\theta$	cotangent
$\csc\theta$	cosecant
D	diffusion constant
d	displacement

Symbol	Definition
d	quark flavor down
dB	decibel
d_i	distance of an image from the center of a lens
d_o	distance of an object from the center of a lens
DC	direct current
E	electric field strength
ε	emf (voltage) or Hall electromotive force
emf	electromotive force
E	energy of a single photon
E	nuclear reaction energy

Symbol	Definition
E	relativistic total energy
E	total energy
E_0	ground state energy for hydrogen
E_0	rest energy
EC	electron capture
E_{cap}	energy stored in a capacitor
Eff	efficiency—the useful work output divided by the energy input
Eff_{C}	Carnot efficiency
E_{in}	energy consumed (food digested in humans)
E_{ind}	energy stored in an inductor

Symbol	**Definition**
E_{out}	energy output
e	emissivity of an object
e^{+}	antielectron or positron
eV	electron volt
F	farad (unit of capacitance, a coulomb per volt)
F	focal point of a lens
$\mathbf{F}$	force
F	magnitude of a force
F	restoring force
F_{B}	buoyant force

Symbol	**Definition**
F_{c}	centripetal force
F_{i}	force input
$\mathbf{F}_{\mathrm{net}}$	net force
F_{o}	force output
FM	frequency modulation
f	focal length
f	frequency
f_0	resonant frequency of a resistance, inductance, and capacitance (RLC) series circuit
f_0	threshold frequency for a particular material (photoelectric effect)

Symbol	Definition
f_1	fundamental
f_2	first overtone
f_3	second overtone
f_B	beat frequency
f_k	magnitude of kinetic friction
f_s	magnitude of static friction
G	gravitational constant
G	green quark color
G	antigreen (magenta) antiquark color

Symbol	**Definition**
g	acceleration due to gravity
g	gluons (carrier particles for strong nuclear force)
h	change in vertical position
h	height above some reference point
h	maximum height of a projectile
h	Planck's constant
hf	photon energy
h_{i}	height of the image
h_{o}	height of the object
I	electric current

Symbol	Definition
I	intensity
I	intensity of a transmitted wave
I	moment of inertia (also called rotational inertia)
I_0	intensity of a polarized wave before passing through a filter
I_{ave}	average intensity for a continuous sinusoidal electromagnetic wave
I_{rms}	average current
J	joule
J/Ψ	Joules/psi meson
K	kelvin
k	Boltzmann constant

Symbol	Definition
k	force constant of a spring
K_α	x rays created when an electron falls into an $n = 1$ shell vacancy from the $n = 3$ shell
K_β	x rays created when an electron falls into an $n = 2$ shell vacancy from the $n = 3$ shell
kcal	kilocalorie
KE	translational kinetic energy
$\text{KE} + \text{PE}$	mechanical energy
KE_e	kinetic energy of an ejected electron
KE_{rel}	relativistic kinetic energy
KE_{rot}	rotational kinetic energy
KE	thermal energy

Symbol	Definition
kg	kilogram (a fundamental SI unit of mass)
$\vec{L}$	angular momentum
L	liter
L	magnitude of angular momentum
L	self-inductance
ℓ	angular momentum quantum number
L_α	x rays created when an electron falls into an $n = 2$ shell from the $n = 3$ shell
L_e	electron total family number
L_μ	muon family total number
L_τ	tau family total number

Symbol	Definition
L_f	heat of fusion
L_f and L_v	latent heat coefficients
$\mathrm{L_{orb}}$	orbital angular momentum
L_s	heat of sublimation
L_v	heat of vaporization
L_z	*z* - component of the angular momentum
M	angular magnification
M	mutual inductance
m	indicates metastable state
m	magnification

Symbol	Definition
m	mass
m	mass of an object as measured by a person at rest relative to the object
m	meter (a fundamental SI unit of length)
m	order of interference
m	overall magnification (product of the individual magnifications)
$m\left({}^{A}\mathrm{X}\right)$	atomic mass of a nuclide
MA	mechanical advantage
m_{e}	magnification of the eyepiece
m_e	mass of the electron
m_ℓ	angular momentum projection quantum number

Symbol	Definition
m_n	mass of a neutron
m_{o}	magnification of the objective lens
mol	mole
m_p	mass of a proton
m_{s}	spin projection quantum number
N	magnitude of the normal force
N	newton
$\mathbf{N}$	normal force
N	number of neutrons
n	index of refraction

Symbol	Definition
n	number of free charges per unit volume
N_{A}	Avogadro's number
N_{r}	Reynolds number
$\mathrm{N \cdot m}$	newton-meter (work-energy unit)
$\mathrm{N \cdot m}$	newtons times meters (SI unit of torque)
OE	other energy
P	power
P	power of a lens
P	pressure
$\mathbf{p}$	momentum

Symbol	**Definition**
p	momentum magnitude
p	relativistic momentum
$\mathbf{p}_{\text{tot}}$	total momentum
$\mathbf{p}'_{\text{tot}}$	total momentum some time later
P_{abs}	absolute pressure
P_{atm}	atmospheric pressure
P_{atm}	standard atmospheric pressure
PE	potential energy
PE_{el}	elastic potential energy
PE_{elec}	electric potential energy

Symbol	**Definition**
PE_s	potential energy of a spring
P_g	gauge pressure
P_in	power consumption or input
P_out	useful power output going into useful work or a desired, form of energy
Q	latent heat
Q	net heat transferred into a system
Q	flow rate—volume per unit time flowing past a point
$+Q$	positive charge
$-Q$	negative charge

Symbol	**Definition**
q	electron charge
q_p	charge of a proton
q	test charge
QF	quality factor
R	activity, the rate of decay
R	radius of curvature of a spherical mirror
R	red quark color
R	antired (cyan) quark color
R	resistance
R	resultant or total displacement

Symbol	**Definition**
R	Rydberg constant
R	universal gas constant
r	distance from pivot point to the point where a force is applied
r	internal resistance
$r_{\perp}$	perpendicular lever arm
r	radius of a nucleus
r	radius of curvature
r	resistivity
r or rad	radiation dose unit
rem	roentgen equivalent man

Symbol	**Definition**
rad	radian
RBE	relative biological effectiveness
RC	resistor and capacitor circuit
rms	root mean square
r_n	radius of the *n*th H-atom orbit
R_p	total resistance of a parallel connection
R_s	total resistance of a series connection
R_s	Schwarzschild radius
S	entropy
$\mathbf{S}$	intrinsic spin (intrinsic angular momentum)

Symbol	**Definition**
S	magnitude of the intrinsic (internal) spin angular momentum
S	shear modulus
S	strangeness quantum number
s	quark flavor strange
s	second (fundamental SI unit of time)
s	spin quantum number
$\mathbf{s}$	total displacement
$\sec\theta$	secant
$\sin\theta$	sine
s_z	*z*-component of spin angular momentum

Symbol	Definition
T	period—time to complete one oscillation
T	temperature
T_{c}	critical temperature—temperature below which a material becomes a superconductor
T	tension
T	tesla (magnetic field strength *B*)
t	quark flavor top or truth
t	time
$t_{1/2}$	half-life—the time in which half of the original nuclei decay
$\tan\theta$	tangent
U	internal energy

Symbol	Definition
u	quark flavor up
u	unified atomic mass unit
$\mathbf{u}$	velocity of an object relative to an observer
$\mathbf{u}'$	velocity relative to another observer
V	electric potential
V	terminal voltage
V	volt (unit)
V	volume
$\mathbf{v}$	relative velocity between two observers
v	speed of light in a material

Symbol	Definition
$\mathbf{v}$	velocity
$\mathbf{v}$	average fluid velocity
$V_{\mathrm{B}} - V_{\mathrm{A}}$	change in potential
$\mathbf{v}_{\mathrm{d}}$	drift velocity
V_{p}	transformer input voltage
V_{rms}	rms voltage
V_{s}	transformer output voltage
$\mathbf{v}_{\mathrm{tot}}$	total velocity
v_{w}	propagation speed of sound or other wave
$\mathbf{v}_{\mathrm{w}}$	wave velocity

Symbol	Definition
W	work
W	net work done by a system
W	watt
w	weight
w_{fl}	weight of the fluid displaced by an object
W_{c}	total work done by all conservative forces
W_{nc}	total work done by all nonconservative forces
W_{out}	useful work output
X	amplitude
X	symbol for an element

Symbol	Definition
${}^Z_A X_N$	notation for a particular nuclide
x	deformation or displacement from equilibrium
x	displacement of a spring from its undeformed position
x	horizontal axis
X_C	capacitive reactance
X_L	inductive reactance
x_rms	root mean square diffusion distance
y	vertical axis
Y	elastic modulus or Young's modulus
Z	atomic number (number of protons in a nucleus)

Symbol	**Definition**
Z	impedance